Behavioural and Molecular Changes Associated with Advanced Paternal Age

Smith, Rebecca

Awarding institution:
King's College London

Behavioural and Molecular Changes Associated with Advanced Paternal Age

Rebecca Georgina Smith

MRC Social, Genetic and Developmental Psychiatry Centre,

Institute of Psychiatry,

King's College, London

A thesis submitted to King's College, London

for the degree of Doctor of Philosophy (PhD) 2012

Thesis Abstract

Epidemiological studies highlight a robust association between advanced paternal age and risk for several neuropsychiatric disorders including autism, schizophrenia and bipolar disorder. In this study a mouse model was used to investigate behavioural and molecular effects of advanced paternal age on the offspring. C57BL/6J females aged 2 months were mated with fathers of the same strain of three different ages: 2 months to represent 'young' fathers, 10 months to represent 'old' fathers and 12 months to represent 'very old' fathers. The offspring were subjected to behavioural tests to investigate their anxiety, locomotion, exploration, memory and social interaction. Significant reductions in exploratory, and particularly social, behaviours in the offspring of old fathers were observed. The deficit in social behaviour is interesting, given that social deficits are a common hallmark across the neuropsychiatric disorders associated with advanced paternal age.

A variety of methods were utilised to investigate any molecular changes that might underlie the observed paternal age effect. Spermatozoa undergo multiple divisions throughout the male lifespan, potentially leading to a higher incidence of *de novo* genomic alterations in the offspring. To investigate whether copy number variation (CNV) was associated with advanced paternal age, DNA from the offspring was subjected to genome-wide CNV analysis using comparative genomic hybridisation (CGH) combined with high-resolution microarrays. No differences in the number or size of CNVs or the genes affected by CNVs between the offspring of young fathers and offspring of old fathers were observed. An alternative hypothesis is that epigenetic alterations may explain the relationship between advanced paternal age and behavioural alterations in offspring. Some regions of the genome, including many imprinted loci and transposable elements escape epigenetic resetting during germ cell development. DNA from multiple tissues was quantified for paternal-age associated differences in global DNA methylation, in addition to DNA methylation across transposable elements and differentially methylated regions (DMRs) associated with brain expressed imprinted genes. An increase in global methylation levels with advanced paternal age were observed, as well as methylation changes at many individual sites in the imprinted DMRs and consistently across DMRs associated with *Gnas*, *Kcnq1ot1* and *Mcts2*. Finally, microarray-based gene expression analysis was carried out on RNA from frontal cortex to look for transcriptomic differences associated with paternal age. Several genes demonstrated

differential expression with advanced paternal age, with enrichment for genes associated with cell signalling, inflammatory response and DNA repair. This study provides strong evidence for deleterious effects of advancing paternal age on social and exploratory behaviour and suggests that incomplete epigenetic reprogramming and altered gene expression are plausible explanatory factors.

Abbreviations

ART	Assisted Reproductive Technology
ASD	Autism Spectrum Disorder
BD	Bipolar Disorder
CGH	Comparative Genomic Hybridisation
CNP	Copy Number Polymorphism
CNV	Copy Number Variation
DAT	Disassemble Array Tool
DMR	Differentially Methylated Region
DNA-PKcs	DNA-Dependent Protein Kinase Catalytic Subunit
DNMTs	DNA Methyltransferases
DSM	Diagnostic and Statistical Manual of Mental Disorders
dNTPs	Deoxynucleoside Triphosphates
DSB	Double Strand Breakage
DZ	Dizygotic
EOBD	Early-Onset Bipolar Disorder
EPM	Elevated Plus Maze
FISH	Fluorescence *In Situ* Hybridization
FoSTeS	Fork Stalling and Template Switching
HATs	Histone Acetyltransferases
HDACs	Histone Deacetylases
IAP	Intracisternal-A Particle
ICR	Imprinting Control Region
IPA	Ingenuity Pathway Analysis
IVF	In-Vitro Fertilization
LINE-1	Long-Interspersed Nuclear Element-1
LTR	Long Terminal Repeat

LUMA	Luminometric Methylation Assay
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight
MBDs	Methyl-CpG-Binding Domain Proteins
MSRE	Methylation Sensitive Restriction Enzyme
MZ	Monozygotic
NAHR	Non-Allelic Homologous Recombination
ncRNAs	Non-Coding RNAs
NHEJ	Non-Homologous End Joining
OR	Odds Ratio
PCA	Principal Component Analysis
PCI	Phenol-Chloroform-Isoamyl Alcohol
PCR	Polymerase Chain Reaction
PGC	Primordial Germ Cell
PK	Proteinase K
PMAT	Precision Mixer Alignment Tool
QC	Quality Control
qPCR	Quantitative PCR
RPM	Raven's Progressive Matrices
RR	Relative Risk
RSN	Robust Spleen Normalization
SAP	Shrimp Alkaline Phosphatise
SD	Standard Deviation
SEM	Standard Error Mean
ssDNA	Single Stranded DNA
TBE	Tris-Borate-EDTA
WGCNA	Weighted Gene Co-Expression Network Analysis
XCI	X-Chromosome Inactivation

Contents Page

List of Figures

List of Tables

Declarations

Breeding of the animals used in this thesis was carried out by Dr Cathy Fernandes.

Social behaviour tasks from Chapter 3 were carried out by Dr Rachel Kember.

Sperm samples used in Chapter 5 were provided by Dr Paul Potter.

Running of Illumina expression arrays in Chapter 6 was carried out by Source Bioscience.

All other laboratory work and analyses were my own work.

Acknowledgements

First and foremost I would like to thank my supervisors. Thank you to Jon Mill for all his help, support, guidance, cheery attitude and infectious enthusiasm over the course of my PhD, and for agreeing to be my supervisor in the first place. I would also like to extend a sincere thank you to Ian Craig for his confidence in me as well as hiring me as an RA in the beginning of my life at the SGDP and providing me with the amazing opportunity to do a PhD. A special thank you goes to Cathy Fernandes for your invaluable help and general awesomeness, not only as a mentor but also as an amazingly generous and kind friend. Another special thank you goes to Leo Schalkwyk for your fabulous stats knowledge, calming attitude and occasionally off the wall hilarity. Also huge thanks to Avi Reichenberg for his help and support over the course of my PhD.

During my time are the SGDP I was lucky enough to have made many friends who have provided me with fun times, emotional support and fond memories. I would like to thank the psychiatric epigenetics group Emma D, Helen, Tim, Yogan, Cybele and Manu for many things including providing much tasty cake. I would especially like to thank Chloe, Rachel, Ruth and Katie who I consider invaluable friends and colleagues. Thank you to past and present office mates, Lucy, Sarah J, Agnes, Ken, Lauren, Alinda, Lourdes and Sophia for putting up with me on a daily bases. Thank you to the lab team Tim, Liz, Charly, Bernie and Gavin for your hard work and help. Thank you to Aaron, Emma M, Laura (Jaeger Bomb), Nick, Hannah, Helena, Rob, Sarah CW, Chris, Zarsh, Matt, Stuart and Mike P for just being great people.

Finally I would like to thank three extremely important people. David, you have been and will hopefully continue to be an amazing friend and boyfriend. I cannot thank you enough for all the support and patience during this time of emotional instability called my PhD. Lastly, an amazingly huge thank you goes to my parents. I can't even begin to say all the things I have to thank you for but mostly your unwavering support and pride. Thank you so much and I love you!

Publications Arising from this Thesis

Chapter 1

Rebecca Smith and Jonathan Mill

Epigenetics and Chronic Diseases: An Overview

Epigenetic Aspects of Chronic Disease, 2011

Chapter 3

Smith RG, Kember RL, Mill J, Fernandes C, Schalkwyk LC, Buxbaum JD, Reichenberg A

Advancing paternal age is associated with deficits in social and exploratory behaviors in the offspring: a mouse model.

PLoS One. 2009 Dec 30; 4(12):e8456

Chapter 5

Smith, RG, Reichenberg A, Kember RL, Buxbaum JD, Schalkwyk LC, Fernandes C, Mill J

Epigenetic Differences at Brain-Expressed Imprinting Loci Are Associated With Advanced Paternal Age

Submitted

Publications from Chapter 4 and 6 to be submitted at a later date

Chapter 1 - Introduction

1.1 Parental Age in the General Population

As the average age at which people are having children increases there are many implications for health and development. Considerable focus is often placed on the trend for increasing maternal age. In North America in 1969, for example, the median age for women to have their first child was 21.3 years old but by 1994 this age had increased to 24.4 years old. During the same period the percentage of women bearing their first child over the age of 30 rose dramatically from 4.1% to 21.2% (Heck, Schoendorf et al. 1997). In Europe the mean age of women giving birth increased 1.5 years between 1980 and 1993 (Breart 1997) and during 1998 and 2008 the percentage of births by women in the United Kingdom over 35 years has increased from 14.5% to 20.1% (Statistics 2009). Due to developments in fertility treatments such as *in vitro* fertilisation, multiple marriages and social attitude changes to women having careers, the age at which many women are having children is increasing.

Men are also fathering children at a later age, in part due to an increase in divorce rates and multiple marriages, and social change. In the period between 1998 and 2008, the percentage of men fathering children over the age of 40 increased from 10% to 16% in the UK (*Figure 1.1*) *(Statistics 2009)*. There has also been a marked increase in men over the age of 50 fathering children (Salonia, Matloob et al. 2011). As the reproductive period for men naturally lasts longer than for women, whose fecundity declines dramatically after 40, the potential for higher paternal age is considerably larger.

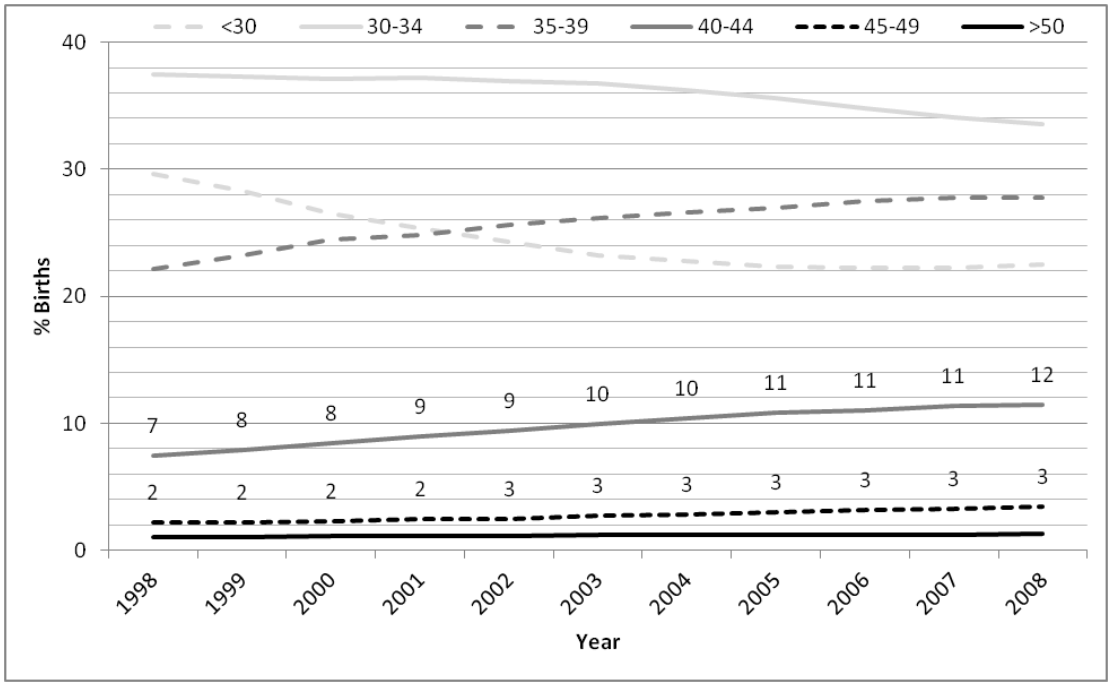


Figure 1.1 - Percentage of UK Births by Father's Age

Adapted from Review of the National Statistician on births and patterns of family building in England and Wales, 2008 (Statistics 2009)

1.2 Parental Age and Disease

1.2.1 Parental Age, Birth Defects and Cancer

The full implications of advancing parental age are not known. Most research has focussed on the consequences of advanced maternal age, which has been associated with many disorders and is strongly linked to negative outcomes related to foetal health and survival. Both advanced maternal and paternal age have been shown to contribute to negative birth outcomes such as low birth weight (Reichman and Teitler 2006) and a maternal age of over 35 years increases the risk of spontaneous abortions. When increased maternal age is combined with a paternal age of over 40, the risk of spontaneous abortions increases further (de la Rochebrochard and Thonneau 2002). Increases in maternal age have been linked to an increased incidence of stillbirth (Huang, Sauve et al. 2008) as well as preeclampsia (Jacobsson, Ladfors et al. 2004). Advanced maternal age has also been shown to influence the risk of a variety of diseases in the offspring. For every 5 year increase in maternal age, a 5% increase in

childhood diabetes has been shown (Cardwell, Stene et al. 2010). Advanced maternal age has also been associated with an increased incidence of childhood cancers, with the greatest effect being for leukaemia and Wilm's tumour (Johnson, Carozza et al. 2009). A maternal age over 40 is associated with an increased incidence of acute lymphoblastic leukaemia (Odds Ratio (OR) 1.94) and acute non-lymphoblastic leukaemia (OR 3.07) in the offspring.

Less publicized is the observation that paternal age has also been shown to increase the risk of disease and physiological conditions in the offspring. A paternal age of over 45 confers substantially increase in risk (Relative Risk (RR) 3.3) for mental retardation of unknown aetiology in the offspring when taking mothers age into account (Zhang 1992). Advanced paternal age also increases the risk of certain types of cancer and a paternal age of over 45 is associated with an increased incidence of retinoblastoma in the offspring resulting from assumed *de novo* germ cell mutations (OR 2.96) (Dockerty, Draper et al. 2001). In addition, numerous birth defects have been associated with advancing paternal age; in a study looking at birth defects in British Columbia, McIntosh *et al* showed that a paternal age greater than 40 was associated with increased risk for birth defects in the offspring. These included neural tube defects (RR 1.6), anencephaly (RR 2.0), spina bifida (RR 1.5), hydrocephaly (RR 2.6), congenital cataracts (RR 1.6), tracheoesophageal fistula (RR 3.9), reduction defect in upper limbs (RR 2.4), reduction defect in lower limbs (RR 3.2) and anomalies of diaphragm (RR 1.8) (McIntosh, Olshan et al. 1995).

1.2.2 Parental Age and Risk for Neurological and Psychiatric Disease

Beyond its influence on birth defects and physiological conditions, there is increasing evidence to suggest that advanced parental age is associated with the risk of developing neurological, behavioural and psychiatric disorders in the offspring. The most notable example is the increased risk of Down's syndrome in the offspring of older mothers. It is widely accepted that as the maternal age increases, the risk of the child suffering from Down's syndrome also goes up due to the higher incidence of chromosome 21 trisomy (Hook 1976; Huether, Ivanovich et al. 1998). The parental age effect is also seen in more complex neurobiological/neuropsychiatric phenotypes. Alzheimer's disease risk, for example, has been shown to be influenced by paternal age, with the offspring of men with a higher paternal age

being more likely to develop pre-senile Alzheimer's disease (Whalley, Thomas et al. 1995) and in a separate study, fathers of Alzheimer's disease patients were significantly older at the child's birth (35.7 years) than the fathers of non-patients (32.6 years) (Bertram, Busch et al. 1998).

Intelligence has previously been shown to have a relationship with both advanced maternal age and paternal age. Parental age shows an inverted *u* shaped relationship with IQ with the offspring of parents younger than 20 and older than 45 showing similar scores. IQ scores for advanced parental age groups steadily decrease as the age of the parents increase (Malaspina, Reichenberg et al. 2005) (*Figure 1.2*). Maternal age has also independently been associated with intelligence, showing an inverted *u* shaped relationship with both a decrease with low maternal age and advanced maternal age (Zybert, Stein et al. 1978; Fergusson and Lynskey 1993). An inverted *u* shaped curve is also observed between paternal age and learning capacity in the offspring. The offspring of fathers aged 17-19 performed worse than fathers 40-44 but both groups showed significantly lower learning capacity than the offspring of fathers 30 years of age (Auroux, Mayaux et al. 1989).

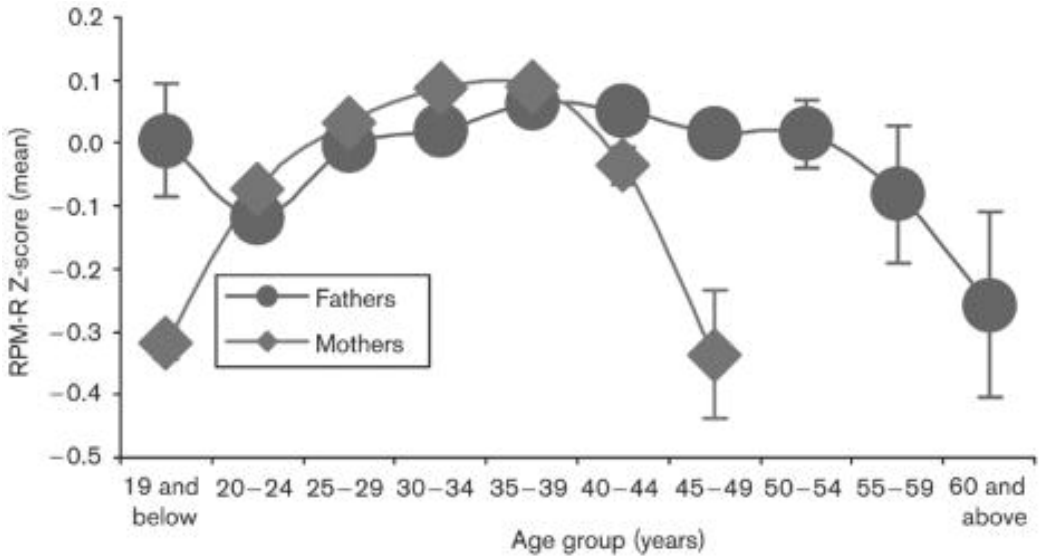


Figure 1.2 - Intelligence and Parental Age

Raven's Progressive Matrices-R (RPM-R) plotted against parental age. From (Malaspina, Reichenberg et al. 2005)

Although there is convincing evidence for the role of parental age in many disorders, some of the strongest evidence relates to the association of advanced parental age with neuropsychiatric disorders, specifically autism spectrum disorders (ASD) (*Table 1.1*), schizophrenia (*Table 1.2*) and bipolar disorder (BD) (*Table 1.3*). Taken together, epidemiological studies provide convincing evidence for the role of parental age in the risk of developing ASDs, schizophrenia and BD.

Table 1.1 - Studies of Parental Age and ASDs

Author	Reference Age	N	PA Finding	MA Finding
(Gillberg 1980)	NA	20 cases	Fathers average age 34 years in cases compared with lower in controls (age not given)	Mean maternal age 31 years compared with 26 years in the controls
(Mouridsen, Rich et al. 1993)	NA	328 cases	Mean age of case fathers was 31.14 years compared 29.68 years in controls	Mean mothers age 27.05 in cases compared to 26.67 in controls
(Croen, Grether et al. 2002)	<20	4356 ASD 3.5mill controls	Not investigated	At ages 30-34, RR = 2.7 At ages >35, RR = 3.4
(Hultman, Sparen et al. 2002)	<30	408 cases 2040 controls	At ages >40, OR = 1.71	At ages >40, OR = 1.84
(Glasson, Bower et al. 2004)	NA	465 cases 481 case sibs 1313 controls	Mean age of case fathers was 31.74 years compared 30.31 years in controls	Mean age of case mothers was 28.62 years compared to 27.01 years in controls
(Lauritsen, Pedersen et al. 2005)	25-29	818 cases 943,664 controls	At ages 40-45, RR = 1.6	No significant association observed
(Maimburg and Vaeth 2006)	25-29	473 cases 4730 controls	No significant association observed	At ages >35, OR = 1.3
(Reichenberg, Gross et al. 2006)	15-29	Population of 378,891	At ages 40-49, OR = 5.75	At ages >40, OR = 2.68
(Cantor, Yoon et al. 2007)	NA	Population of 2,492,308	30-39 54.7% in ASD compared to 41.9% in controls	Not investigated
(Croen, Najjar et al. 2007)	25-29	593 cases 132,251 controls	At ages 35-39, RR = 1.38 At ages >40, RR 1.52	At ages 35-39, RR = 1.2 At ages >40, RR = 1.27
(Durkin, Maenner et al. 2008)	25-29	1,251 cases 253,347 controls	At ages >40, OR = 1.6	At ages >40, OR = 1.4

(Tsuchiya, Matsumoto et al. 2008)	<29	84 cases 208 controls	At ages >33, OR = 3.09 10-year increase in paternal age OR = 2.54	No significant association observed
(Grether, Anderson et al. 2009)	NA	23,311 cases 7.5mill controls	10-year increase in paternal age OR = 1.22	10 year increase in maternal age, OR = 1.38
(Gabis, Raz et al. 2010)	NA	268 cases 1.2mill controls	At ages 35-44, 39% in ASD compared to 28% in controls At ages >45, 5% in ASD compared to 3.5% in controls	Not investigated
(Golding, Steer et al. 2010)	NA	86 cases 13,885 controls	No significant association observed	10 year increase in maternal age, OR = 1.6 10 year increase in grandmaternal age, OR = 1.66
(Hultman, Sandin et al. 2010)	<29	883 cases 1.1 mill controls	At ages >50, OR = 2.2 At ages >55, OR = 4.36	No significant association observed
(Lundstrom, Haworth et al. 2010)	25-34	Twins Sweden: 11,122 UK: 13,524	Sweden: At ages 45-50, OR = 1.9 Sweden: At ages >51, OR = 3.24 UK: At ages 45-50, OR1.66 UK: At ages >51, OR = 3.59	No significant association observed
(Sasanfar, Haddad et al. 2010)	25-29	179 cases 1611 controls	At ages >40, OR = 2.03	No significant association observed
(Shelton, Tancredi et al. 2010)	<25	12,159 cases 4.9mill controls	At ages >40, OR = 1.78	At ages >40, OR = 1.77
(Zhang, Lv et al. 2010)	<30	95 cases 95 controls	At ages >30, OR = 2.62	No significant association observed
(Buizer-Voskamp, Laan et al. 2011)	25-29	14,231 cases 56,924 controls	At ages >40, OR = 1.23	No significant association observed
(Krishnaswamy, Subramaniam et al. 2011)	20-29	1972 respondants	At ages >50, OR = 2.38	No significant association observed

Table 1.2 - Studies of Parental Age and Schizophrenia

Author	Reference Age	N	PA Finding	MA Finding
(Hare and Moran 1979)	NA	8000 cases	Mean paternal age 34 years compared with 32 years in the controls	Mean maternal age 31 years compared with 29 years in the controls
(Gillberg 1982)	NA	155 cases	Average age of fathers older in cases than in controls	60% of mothers >30 compared with 27% of children in controls
(Kinnell 1983)	NA	342 cases 1,817 controls	Mean paternal age 34 years compared with 32 years in the controls	Mean maternal age 30 years compared with 29 years in the controls
(Malaspina, Harlap et al. 2001)	20-24	658 cases 87,907 controls	At ages 45-49, OR = 2.02 At ages >50, OR = 2.96	No significant association observed
(Brown, Schaefer et al. 2002)	15-24	Population of 19,044	At ages >45 Rate ratio 3.6	Not investigated
(Byrne, Agerbo et al. 2003)	20-24	7704 cases 192,590 controls	At ages >55, OR = 2.42	At ages >45, OR = 1.43
(Zammit, Allebeck et al. 2003)	15-24	Population of 50,087	At ages 45-54, OR = 1.6 At ages >55, OR = 3.8 10-year increase in paternal age, OR = 1.3	No significant association observed
(El-Saadi, Pedersen et al. 2004)	20-24	117 cases 146 controls	Sweden: At ages >35, OR = 2.42 Denmark: At ages 50-54, OR = 1.62 Denmark: At ages >55, OR = 1.84	Denmark: At ages >50, OR = 1.35
(Sipos, Rasmussen et al. 2004)	21-24	Population of 754,330	At ages >50, HR = 4.62 10-year increase in paternal age, HR = 1.47	No significant association observed
(Tsuchiya, Takagai et al. 2005)	<28	99 cases 381 controls	At ages 29-31, OR = 2.08 At ages >32, OR = 3.0	No significant association observed
(Torrey, Buka et al. 2009)	<30	168 cases 25,025 controls	At ages >35, OR = 1.35	No significant association observed

(Petersen, Mortensen et al. 2010)	25-29	Population of 2.2mill	At ages >55, RR = 1.93 10-year increase in paternal age RR 1.15	No significant association observed
(Buizer-Voskamp, Laan et al. 2011)	25-29	14,231 cases 56,924 controls	35-39 OR 1.24 >40 OR 1.27	No significant association observed
(Lee, Malaspina et al. 2011)	NA	259 cases	Paternal age of 41 years higher than in controls	Mean maternal age 33 years higher than average
(Miller, Messias et al. 2010)	25-29	14,568 cases 3mill controls	45-49 OR 1.24 >50 OR 1.79	Not investigated
(Naserbakht, Ahmadkhaniha et al. 2011)	NA	220 cases 220 controls	34.5% of case fathers were over 32 years old compared to 15% in controls	11.5% of case mothers were over 32 years old compared to 6.5% in controls

Table 1.3 - Studies of Parental Age and BD

Author	**Reference Age**	**N**	**PA Finding**	**MA Finding**
(Frans, Sandin et al. 2008)	20-24	13,428 cases 7.3mill controls	At ages 40-44, OR = 1.96 At ages 45-49, OR = 2.32 At ages >50, OR = 2.63	No significant association observed
(Menezes, Lewis et al. 2010)	21-24	Population of 711,989	10 year increase in paternal age, OR = 1.2 At ages 40-44, OR = 1.85	10 year increase in maternal age, OR = 1.23

The association between advanced paternal age and risk of developing neuropsychiatric disorders is well established. At least nineteen published studies report an association between advanced paternal age and ASD. Some studies report a linear association with paternal age (Tsuchiya, Matsumoto et al. 2008; Grether, Anderson et al. 2009) with higher paternal age increasing the risk of developing ASD. However, most studies report that once a certain paternal age threshold is reached, then risk of offspring-ASD increases dramatically before reaching a plateau. Most studies report that the critical age is approximately 40 years old and that this risk increases further as paternal age exceeds 50. A potential problem with studies of a paternal age greater than 50 is that most are characterised by a very small number of cases with fathers in this age category, although this number is generally larger than observed in control samples (Gabis, Raz et al. 2010). Furthermore, the average age of fathers in cases is consistently estimated to be 1-2 years older than in controls (Mouridsen, Rich et al. 1993; Glasson, Bower et al. 2004).

Similarly, at least sixteen published studies report an association between advanced paternal age and the risk of developing schizophrenia (or other psychotic disorders). Some studies again report a linear effect, with an increase in risk with every 10-year increase in paternal age (Zammit, Allebeck et al. 2003; Sipos, Rasmussen et al. 2004; Petersen, Mortensen et al. 2010), and overall the average age of fathers in cases is two years older than in controls (Hare and Moran 1979; Kinnell 1983). Unfortunately, because different epidemiological studies have chosen different thresholds for 'advanced' paternal age, it is difficult to define at what age the effect starts. However, there is a consistent observation that a paternal age of over 45 leads to an increase risk of developing schizophrenia. Some studies report a further increase in risk when paternal age increases to over 50 (Malaspina, Harlap et al. 2001; Zammit, Allebeck et al. 2003; El-Saadi, Pedersen et al. 2004; Sipos, Rasmussen et al. 2004; Miller, Messias et al. 2010), but again the numbers in this group are low.

Two studies report an association between advanced parental age and BD (Frans, Sandin et al. 2008; Menezes, Lewis et al. 2010). Both studies find that a paternal age of over 40 is associated with an increased risk of developing BD, and one found that this risk continued to increase as paternal age increased (Frans, Sandin et al. 2008).

The studies reviewed above indicate that paternal age contributes to the risk of developing a number of psychiatric disorders including ASD, schizophrenia and BD. However, it is not just risk of developing these disorders that is affected. For example, looking at paternal age in ASD, Lundström and colleagues observed that twin concordance rates also varied as a function of paternal age. As the age of the father increases, the twin concordance rate between both monozygotic (MZ) and dizygotic twins (DZ) increases (Lundstrom, Haworth et al. 2010) (*Figure 1.3*). The authors report an MZ twin concordance rate approaching 100% in two separate cohorts of patients when paternal age was over 40 years. The implications of this are that the increasing paternal age is making MZ twins more similar, potentially caused by elements in the zygote before it splits, or in the gametes which create the zygote. As the same sperm and oocyte produce both members of an MZ twin-pair, the paternal-age associated disease-causing factors could result from information contained in the gametes or the quality of gametes, which as will be discussed in section 1.4.1, declines with increasing age. In DZ twin pairs, the concordance rate increased from 35% with fathers of 25-35 to 70% in fathers over 40 years (Lundstrom, Haworth et al. 2010), implying that defects in sperm which cause autism (or other disorders) become more common with advancing male age thus leading to a greater chance of the development of psychiatric disorders.

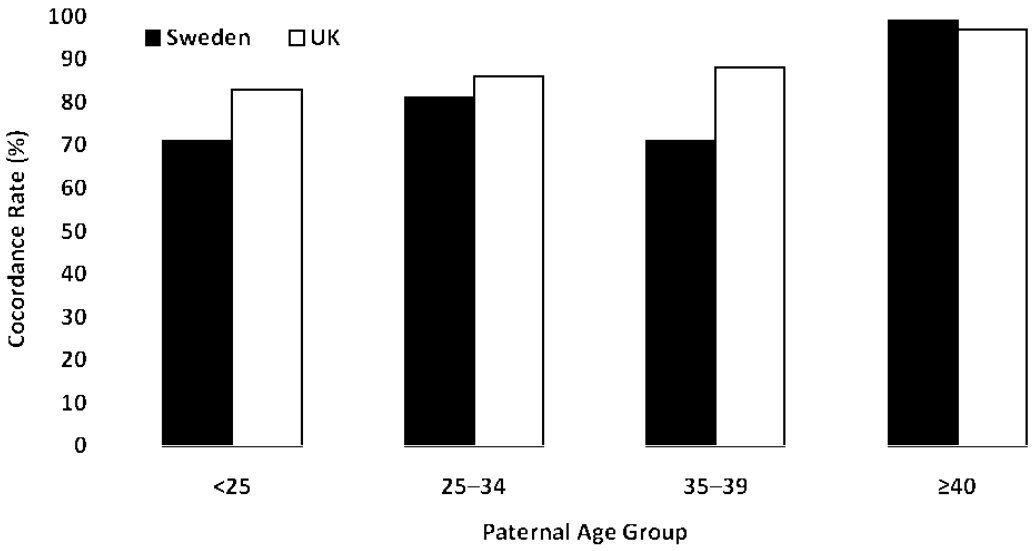


Figure 1.3 – ASD MZ Twin Concordance Rates with Paternal Age

Adapted from (Lundstrom, Haworth et al. 2010)

Sex-ratios in ASD are also affected by advanced paternal age. The typical sex ratio for ASD is heavily male-biased, with 4:1 ratio of males to females (M:F) (Fombonne 2003). A paternal age related shift in sex ratios was first shown in a study that found a M:F ASD ratio in the offspring of fathers less than 40 years old of 5.6:1 which dropped to 2.3:1 in the offspring of fathers over 40 (Reichenberg, Gross et al. 2006). Anello and colleagues also showed that the sex ratio varies dramatically as a function of paternal age: the offspring of fathers of less than 30 years had a ratio of 6.1:1, but this dropped significantly to 1.21:1 with a paternal age of over 45 (Anello, Reichenberg et al. 2009). Interestingly, the same effect was not seen with maternal age in this study. Paternal age has also been shown to affect the overall sex ratio of live births in Caucasian populations, with female offspring being more common in older fathers than male offspring (Ruder 1985). Again, these observations hint at potential biological mechanisms behind the paternal age effect, and suggest that the manifestation of ASD in the offspring of older fathers may be aetiologically distinct to that seen in the offspring of younger fathers where sex-specific factors may be more important.

As maternal and paternal age are highly correlated in human populations, identifying independent maternal- and paternal-age effects is difficult and it is often not clear whether the predominant effect comes from paternal age, maternal age or a combination of both. Some of the studies reviewed in this chapter show an advanced maternal age effect as well as a paternal age effect (*Table 1.1*, *Table 1.2* and *Table 1.3*), with some showing a higher mean age in cases compared to controls and others having an increased risk with increasing maternal age. Although the opportunity for males to sire children at a later age is higher than in females for whom the reproductive lifespan is biologically constrained, in human epidemiological studies it is not possible to fully elucidate the effect of paternal age independently from maternal age and vice versa. Human epidemiological studies are also potentially confounded by genetic variation and factors in the environment, and we can only draw speculative conclusions about what is truly causing the apparent increased prevalence of neuropsychiatric disorders in the offspring of older parents. It has been argued that the use of animal models of parental age may be a more powerful tool to disentangle these effects and investigate mechanism(s) behind their action (Mill 2007) (see section 1.7). In particular, the use of inbred mouse strains, bred for multiple generations to ensure homozygosity, enables researchers to control for both between-individual genetic variation and the environment. The primary focus of this thesis, therefore, is the development of an animal model of

advanced paternal age, with subsequent behavioural and molecular analyses, in order to better understand the human epidemiological findings of a link between older fathers and neuropsychiatric disorders such as ASD and schizophrenia.

1.3 Clinical Features of ASD, Schizophrenia and BD: Overlap in Social Deficits

ASDs consist of a combination of disorders including autism, Rett syndrome, and Asperger syndrome. All ASDs are characterised in the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV (American-Psychiatric-Association 2000) by multiple symptoms, but are mainly defined by impairments or delays in social functioning (Kanner 1968), impairments in non-verbal behaviours such as eye contact and facial expression, impaired language and communication, as well as repetitive behaviours (Kanner 1968) or a narrow sometimes obsessive range of interests (American-Psychiatric-Association 2000) and developmental delays in social interaction and imaginative play. According to the DSM IV, schizophrenia is diagnosed when a patient displays two or more of symptoms of delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behaviour and negative symptoms, such as affective flattening, lack of additional, unprompted content in speech, a general lack of desire, drive, or motivation to pursue meaningful goals. It is also characterised by social impairments and being socially withdrawn (Mueser, Bellack et al. 1991). Bipolar disorder (BD) is characterised in the DSM IV by one or more manic episodes and episodes of depression. Bipolar patients also display social phenotypes of shyness or social anxieties (Simon, Otto et al. 2004).

One hypothesis about why there is a similar paternal age effect common to multiple neuropsychiatric disorders (i.e. ASD, schizophrenia and BD) is that paternal age mediates a behavioural phenotype common to all three disorders (i.e. social behavioural deficits) rather than being independently associated with each disorder. Evidence for this hypothesis comes from studies examining social functioning traits in non-disease cohorts, which conclude that with increasing paternal age, the chance of the offspring having impaired social functioning increases (Weiser, Reichenberg et al. 2008) (*Figure 1.4*). Compared to the reference age of 25-39, a paternal age of over 45 had an OR of 1.52 for social deficits. Of note, the same effect was

not observed with advanced maternal age. In a similar study into autism-like traits associated with paternal age, a *u* shaped relationship in social scores was observed in two independent samples of twins with autism with both the offspring of old and young fathers scoring highly (Lundstrom, Haworth et al. 2010).

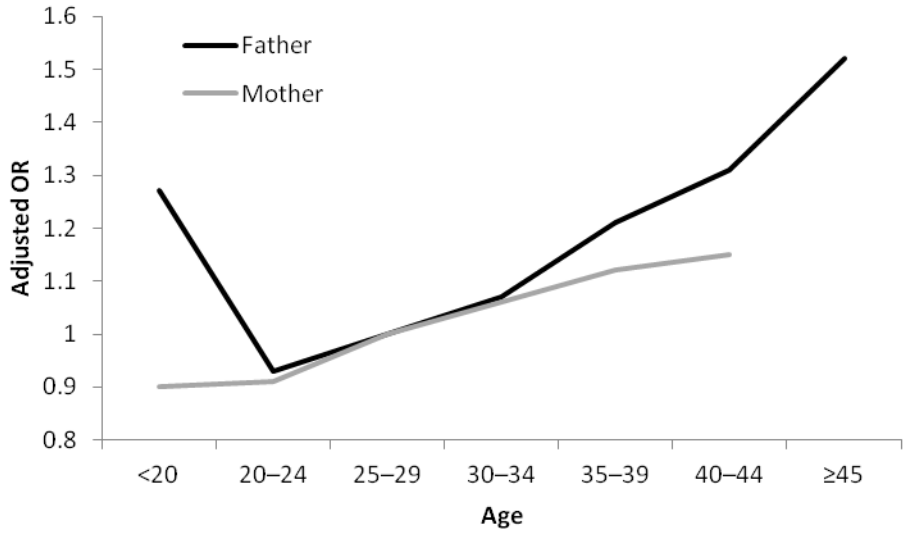


Figure 1.4 – Parental Age and Risk of Social Functioning Deficits in Male Adolescents

OR of impairment in social functioning plotted against parental age. Adapted from (Weiser, Reichenberg et al. 2008).

1.4 Potential Mechanisms of Parental Age Effects

The potential mechanism by which the paternal age effect is mediated has yet to be elucidated although there are many hypotheses, many of which are explored in the course of this thesis. In human males, it is estimated that there are 30 cell divisions of germ-cells before puberty then a further 23 divisions every year. There are also a further four mitotic and two meiotic divisions before the sperm are fully formed. This equates to 610 chromosomal replications by the age of 40 and 840 replications by age 50 (Crow 2000). In contrast, female germ-cells only undergo 24 divisions in their entire development from primordial germ-cells to mature cells regardless of the woman's age. This includes 22 mitotic divisions, which occur in the last 5 months of gestation up until birth, then 2 meiotic divisions from the age of sexual

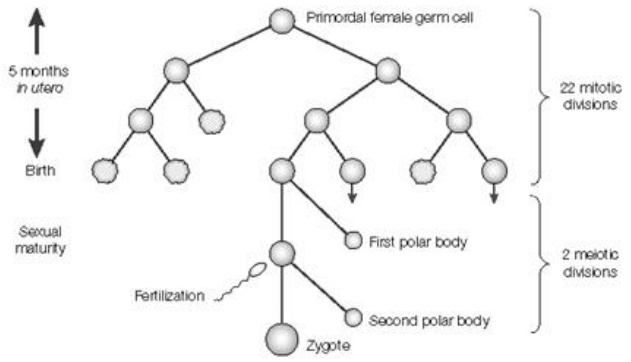
maturity to create the first polar body (*Figure 1.5*). This difference in the number of cell replications undertaken before mature germ-cells are produced suggests potential differences in the mechanisms by which parental age effects become manifest in males and females. Some mechanisms by which the paternal age effect, the focus of this thesis, could be mediated are discussed briefly below.

Figure 1.5 - Germ Cell Development in Humans

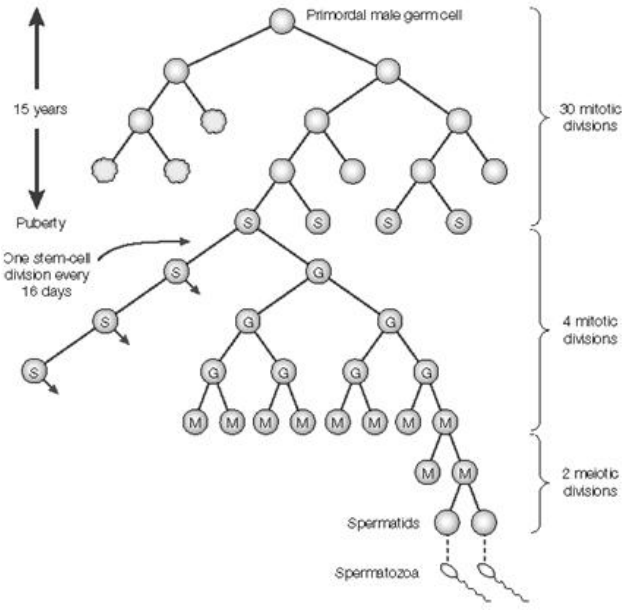
Cell divisions and stages to the development of mature gametes (Pitman 2001)

a) *Female gametes undergo 22 mitotic and 2 meiotic divisions to mature oocytes*
b) *Male gametes undergo 30 mitotic divisions before puberty then one stem-cell divisions every 16 days. A further 4 mitotic and 2 meiotic divisions occur to create mature spermatozoa*

a)



b)



1.4.1 Changes in Sperm Quality with Advancing Male Age

It had been observed that sperm quality declines with increasing male age. Advanced male age was found to negatively correlate with seminal volume (Eskenazi, Wyrobek et al. 2003; Levitas, Lunenfeld et al. 2007; Cardona Maya, Berdugo et al. 2009; Hammiche, Laven et al. 2010), rapid progressive motility (Centola and Eberly 1999; Jung, Schuppe et al. 2002; Eskenazi, Wyrobek et al. 2003; Cardona Maya, Berdugo et al. 2009), percent of motile spermatozoa (Centola and Eberly 1999; Eskenazi, Wyrobek et al. 2003; Wyrobek, Eskenazi et al. 2006; Levitas, Lunenfeld et al. 2007), spermatozoa concentration (Jung, Schuppe et al. 2002; Levitas, Lunenfeld et al. 2007; Cardona Maya, Berdugo et al. 2009), percentage of spermatozoa with tail defects (Centola and Eberly 1999; Jung, Schuppe et al. 2002), percentage with head defects (Jung, Schuppe et al. 2002), the pH of seminal fluid (Jung, Schuppe et al. 2002), total sperm count (Centola and Eberly 1999; Jung, Schuppe et al. 2002; Eskenazi, Wyrobek et al. 2003; Levitas, Lunenfeld et al. 2007; Cardona Maya, Berdugo et al. 2009) and most interestingly for genetic research, positively correlated with a higher percentage in the sperm DNA fragmentation index or sperm chromatin structure analysis (Sartorelli, Mazzucatto et al. 2001; Wyrobek, Eskenazi et al. 2006; Hammiche, Laven et al. 2010). The sperm DNA fragmentation index is a ratio expressed as a percentage of the sperm DNA with both unfragmented dsDNA and fragmented ssDNA (Evenson, Jost et al. 1999). This decline in semen and sperm quality could potentially be the cause of transmission of the paternal age effect through the occurrence of genomic alterations such as *de novo* copy number variations (CNVs).

1.5 De Novo Copy Number Variations

CNVs are defined as deletions or duplications between 1Kb and 5Mb in size (Sebat, Lakshmi et al. 2004) and are a cause of major structural variation in the genome. There are three main mechanisms identified for the creation of CNVs: non-allelic homologous recombination (NAHR), non-homologous end joining (NHEJ) and fork stalling and template switching (FoSTeS), and these are discussed in detail in the introductory section to Chapter 4.

In 2004, two landmark studies showed that CNVs were more common in the human genome than previously thought. Lafrate et al identified 255 loci across the human genome

containing genomic imbalances among unrelated individuals. Twenty-four variants were found to be present in over 10% of the individuals that were examined. Half of these regions were found to overlap with genes, and many coincided with segmental duplications or gaps in the human genome assembly (Iafrate, Feuk et al. 2004). Sebat et al showed that large-scale copy number polymorphisms (CNPs) (CNVs 100kbs and greater) contribute substantially to genomic variation between normal humans. Using comparative genomic hybridisation (CGH) microarray analysis they revealed a total of 221 copy number differences representing 76 unique CNPs containing genes involved in a number of pathways and functions including neurological function and genes known to be associated with disease (Sebat, Lakshmi et al. 2004).

CNVs account for more between-individual base pair variation than any other form of mutation, and have likely played an important role during evolution. Cheng et al presented a global comparison of differences in content of segmental duplication between human and chimpanzee. They determined that 33% of human duplications are not duplicated in chimpanzee, including some which are known to cause disease in humans and estimated a genomic duplication rate of 4-5Mb per million years since divergence. Importantly, the changes identified were shown to result in gene expression differences between the humans and chimpanzees (Cheng, Ventura et al. 2005).

Rare CNVs of large effect, often occurring *de novo*, have been implicated in a wide-range of complex neuropsychiatric disorders including disorders associated with advanced paternal age such as ASD (Marshall, Noor et al. 2008; Glessner, Wang et al. 2009; Fernandez, Roberts et al. 2010), schizophrenia (Stefansson, Rujescu et al. 2008; St Clair 2009; Ingason, Rujescu et al. 2011) and BD (Zhang, Cheng et al. 2009). Autism patients have previously been shown to have a higher incidence of *de novo* CNVs (10%) than in controls (1%) and 1st degree unaffected relatives (Sebat, Lakshmi et al. 2007). In schizophrenia it has been observed that novel CNVs are present in a significantly higher percentage of cases than controls (International Schizophrenia Consortium 2008; Walsh, McClellan et al. 2008; Xu, Roos et al. 2008), and higher still in early-onset cases (Walsh, McClellan et al. 2008). The occurrence of CNVs in cases has been shown to be 1.15 fold higher in cases than in controls that show on average 0.99 CNVs per person (Stone, O'Donovan et al. 2008). BD has also been associated with an

increased burden of CNVs, especially in cases of early-onset. Zhang *et al* showed that singleton CNVs that occur only once in a dataset were seen in 12.3% of controls compared with 16.2% of bipolar patients that had a first incidence of mania below the age of 18 (Zhang, Cheng et al. 2009). Also it has been shown that BD patients have an increased overall burden of CNVs than in controls and higher between early-onset cases and late-onset cases (Priebe, Degenhardt et al. 2011). Together, these studies suggest that CNV burden could mediate the paternal age effect where individuals that show a higher incidence of *de novo* CNVs are subject to gene dosage problems or deletions of coding regions of genes. CNVs located in candidate genomic regions were observed in 10-20% of observed autism cases in two separate studies (Jacquemont, Sanlaville et al. 2006; Sebat, Lakshmi et al. 2007) but consistent CNVs affecting candidate genes have been speculated to account for an even higher percentage (Beaudet 2007). The role of CNVs in the mediation of the paternal age effect will be investigated using a mouse model in Chapter 4.

1.6 Epigenetics

An alternative hypothesis is that the paternal age effect is mediated by epigenetic (literally “above” genetics from the Greek “epi”) changes to the genome that are potentially transmitted via the male germline. The term ‘epigenetics’ was first coined by the British biologist Conrad Waddington in 1942 to describe the underlying developmental mechanisms by which a single genotype produces the diverse range of cellular phenotypes observed in an organism (Waddington 1942). More recent definitions see epigenetics as the study of reversible changes in gene function that occur independently of changes in DNA sequence (Riggs, Martienssen et al. 1996). Epigenetic processes are essential for normal cellular development and differentiation, and allow the long-term regulation of gene function through non-mutagenic mechanisms (Henikoff and Matzke 1997). Once established, epigenetic marks can be stably inherited through mitosis down cell lineages although, as will be described in more detail below and in Chapter 5, they can also be dynamically regulated in response to developmental, environmental, and stochastic factors. A glossary of epigenetic terms is found in *Table 1.4* (taken from (Smith and Mill 2011).

Table 1.4 - Glossary of Epigenetic Terms

Term	Definition	Key Reference(s)
Epigenetics	The heritable, but reversible, regulation of various genomic functions that occur independently of the DNA sequence. Epigenetic regulation is primarily mediated by DNA methylation, physical changes to chromatin structure, and the action of post-transcriptional regulators.	(Henikoff and Matzke 1997)
Chromatin	The complex of DNA, histones, and other proteins that make up chromosomes. Chemical modifications to both DNA and histone proteins are important in regulating the structure of chromatin. Condensed chromatin (heterochromatin), in which the DNA and histone proteins are tightly packed, acts to block the access of transcription factors and other instigators of gene expression. Open chromatin (euchromatin) allows transcriptional factors to access DNA and drive transcription.	(Berger 2007)
DNA methylation	The addition of a methyl group at position 5 of the cytosine pyrimidine ring in CpG dinucleotides in a reaction catalyzed by DNA methyltransferases. DNA methylation disrupts the binding of transcription factors and attracts methyl-binding proteins that are associated with gene silencing and chromatin compaction.	(Jaenisch and Bird 2003)
Histone Modifications	Covalent post-translational histone modifications that occur at specific residues include acetylation, methylation, phosphorylation, SUMOylation, and ubiquitylation. These modulate gene expression via alterations in chromatin structure. Like DNA methylation, histone modifications are highly dynamic and actively regulated by a host of catalytic enzymes.	(Berger 2007)
Genomic Imprinting	An epigenetic process that alters the expression of genes in a parent of-origin specific manner. Genomic imprinting is fundamental to normal mammalian development and growth.	(Davies, Isles et al. 2005)
X-inactivation	X-chromosome inactivation silences genes on one X-chromosome in females to ensure dosage compensation with males via a process involving hypermethylation of CpG islands. X-inactivation in any given cell is typically random and, once established, is maintained so that the inactivated allele is transcriptionally silenced for the lifetime of that cell.	(Avner and Heard 2001)
Epigenetic Inheritance	Epigenetic signals are transmitted mitotically through cell lineages, but are generally assumed to be reset during gametogenesis and thus not transmitted meiotically. Evidence is mounting, however, that the epigenetic marks of at least some mammalian genes are not fully erased during meiosis and may therefore constitute the mechanism for transgenerational epigenetic inheritance.	(Richards 2006)

The most studied and best understood epigenetic modification to the genome is DNA methylation, and this is the focus of the experiments presented in Chapter 5 of this thesis. Other epigenetic modifications include the numerous post-translational modifications (including acetylation, methylation, ubiquitylation, phosphorylation and sumoylation) to the N- termini tails of the core histone proteins around which DNA is wrapped to form nucleosomes, and small non-coding RNAs (ncRNAs), which are believed to underpin chromatin remodelling and be important in epigenetic memory (Mattick and Makunin 2006). Like DNA methylation, histone modifications are highly dynamic and actively regulated by a host of catalytic enzymes, such as histone acetyltransferases (HATs) and histone deacetylases (HDACs), which add and remove acetyl groups respectively (Saha and Pahan 2006). While often investigated independently, epigenetic modifications to DNA and histones are not mutually exclusive, and clearly interact in a number of ways; it is apparent that the classification of epigenetic mechanisms in terms of either gene activation or suppression is too simplistic (Berger 2007). The methyl-binding protein MeCP2, for example, binds specifically to methylated cytosines, attracting histone deacetylases which deacetylate histones (Robertson and Wolffe 2000), and a recent study has shown that histone H3 residues unmethylated at the lysine 4 position recruit DNA methyltransferases resulting in *de novo* DNA methylation (Ooi, Qiu et al. 2007).

DNA methylation refers to the addition of a methyl group to the 5^{th} carbon position of the cytosine pyrimidine ring (*Figure 1.6*). DNA methyltransferases (DNMTs) are the family of enzymes which catalyze the methylation of DNA (Kim, Samaranayake et al. 2009) and to date, four have been identified (Weber and Schubeler 2007). These are DNMT1 which maintains DNA methylation during replication by copying the DNA methylation of the parental DNA strand onto the newly synthesized strand (Leonhardt, Page et al. 1992), DNMT2 which has a weak ability to methylate DNA and is more likely an RNA methyltransferase (Goll, Kirpekar et al. 2006) and DNMT3a and DNMT3b, which are responsible for DNA methylation occurring *de novo* (Okano, Bell et al. 1999). DNMT3a and DNMT3b also work with DNMT1 to maintain the methylation pattern of DNA through cell divisions (Liang, Chan et al. 2002).

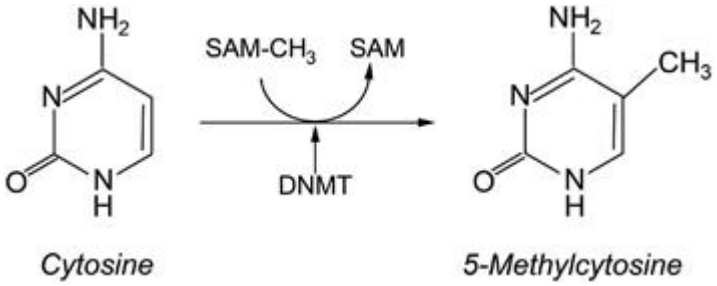


Figure 1.6 - DNA Methylation: Cytosine to 5-Methylcytosine

DNA methylation occurs primarily at CpG dinucleotides, although recent studies have revealed a noticeable amount of non CpG DNA methylation in embryonic stem cells (Lister, Pelizzola et al. 2009). CpG dinucleotides occur at a lower frequency in mammalian genomes than expected, only accounting for ~1% of the human genetic sequence (Lander, Linton et al. 2001) because methylated cytosines are liable to spontaneous deamination to thymine residues (Coulondre, Miller et al. 1978; Bird 1980). The majority of CpG sites in the mammalian genome (70-80%) are methylated; with the exception of those located in CpG-rich regions termed 'CpG islands', which are often associated with the promoter regions of constitutively expressed housekeeping genes (Gardiner-Garden and Frommer 1987). The definition of a CpG island is somewhat arbitrary; they were originally defined as a region at least 200bp in length, a GC content over 50% and an observed/expected CpG ratio of greater than 60% (Gardiner-Garden and Frommer 1987). This was later redefined with more stringent criteria as a region at least 500bp in length, a GC content greater than 55% and an observed/expected CpG ratio that is greater than 65% to lessen the overlap of CpG islands with repetitive elements (Takai and Jones 2002), although many databases such as the UCSC genome browser still use the earlier criteria (Kent, Sugnet et al. 2002). CpG islands roughly cover 40% of all mammalian gene promoters (Larsen, Gundersen et al. 1992) and are generally hypomethylated. DNA methylation at these CpG islands has been the predominant focus of epigenetic research and is associated with chromatin remodelling and reduced gene expression at proximal genes (Jaenisch and Bird 2003). It is increasingly apparent, however, that DNA methylation at CpGs occurring outside promoter CpG islands (~70% of total CpGs), also has an important role in regulating genomic function, particularly during development (Shen, Chow et al. 2006). For example, recent genome-wide analyses have highlighted the importance of non-CpG island promoters (Han, Cortez et al. 2011), intragenic CpG islands

(Medvedeva, Fridman et al. 2010), and CpG island 'shores', which are CpG rich regions up to 2kb away from CpG islands (Irizarry, Ladd-Acosta et al. 2009).

DNA methylation provides a dynamic mechanism influencing the developmental regulation of gene expression (Fang, Wang et al. 2003; Xiang, Zhao et al. 2008). For example, gene expression is often negatively correlated with DNA methylation across promoter CpG islands because methylated DNA prevents the binding of certain transcription factors (Bird and Wolffe 1999). Furthermore, highly methylated regions bind methyl-CpG-binding domain proteins (MBDs), which can act as adaptors between methylated DNA and histone-modifying enzymes and chromatin remodelling proteins. MECPs recruit histone-modifying enzymes to stretches of methylated DNA (Hendrich and Tweedie 2003). The resulting closed and compacted chromatin structure (heterochromatin) blocks the binding of transcription factors and thus represses gene transcription (Jenuwein and Allis 2001). Inversely, low DNA methylation at gene promoters leads to an open chromatin structure (euchromatin) and is often associated with higher levels of gene expression (*Figure 1.7*). A classic example of the silencing role of DNA methylation is seen in the agouti (A^{vy}) mouse model. A reduction in DNA methylation at an intracisternal-A particle (IAP) upstream of the Agouti gene leads to a change in coat colour from brown to yellow by increasing transcription from a cryptic promoter in the gene (Dolinoy, Huang et al. 2007).

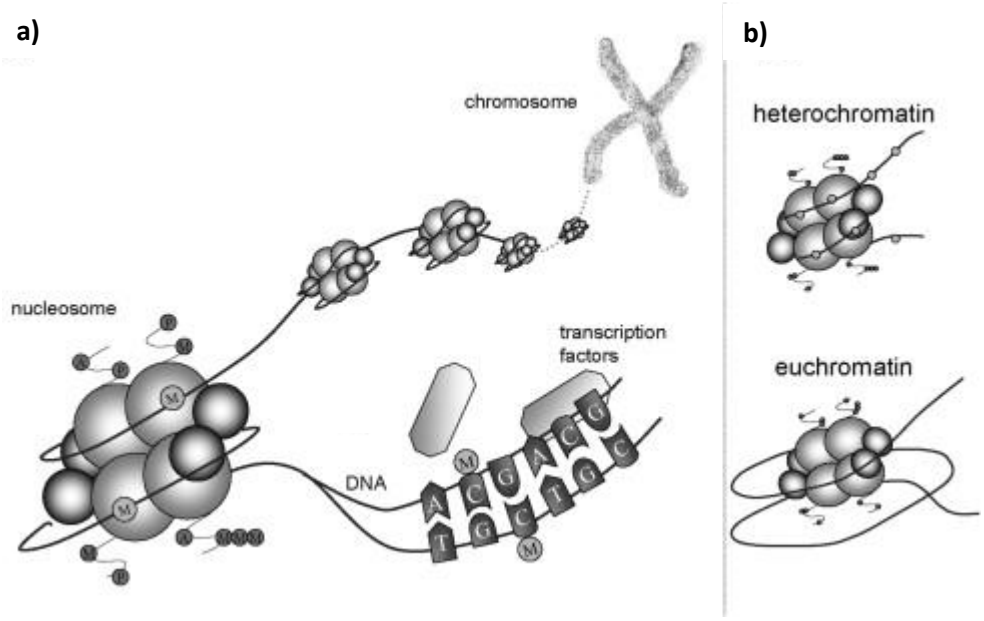


Figure 1.7 - Epigenetic Modifications and Chromatin Structure

(Pidsley and Mill 2011)

a) *DNA is wrapped around a histone core, forming a nucleosome. DNA methylation at CpG dinucleotides and covalent modifications to histone tails alter the physical structure of the chromosome*

b) *Heterochromatin (closed chromatin structure) and euchromatin (open chromatin structure)*

1.6.1 Epigenetic Changes in Complex Neuropsychiatric Disease

Epigenetic processes in the brain have been associated with a range of biological and cognitive processes including neurogenesis (Ma, Marchetto et al. 2010), brain growth and development (Pidsley, Dempster et al. 2010), learning and memory (Lubin, Roth et al. 2008), drug addiction (Renthal and Nestler 2009), neurodegeneration (Migliore and Coppede 2009), and circadian rhythm (Nakahata, Grimaldi et al. 2007). It has been widely speculated, therefore, that epigenetic dysfunction in the central nervous system may be involved in a spectrum of psychiatric disorders including those epidemiologically-linked with advanced paternal age. DNA methylation differences at specific loci have been observed in autism (Nguyen, Rauch et al. 2010), schizophrenia (Petronis, Gottesman et al. 2003), bipolar disorder

(Carrard, Salzmann et al. 2011) and major psychosis (Mill, Tang et al. 2008; Dempster, Pidsley et al. 2011). The role of DNA methylation in psychiatric conditions is further reviewed in section 5.2.3.

1.6.2 Genomic Imprinting

Across the majority of the mammalian genome, DNA methylation is assumed to be complementary on both alleles, although there are several exceptions where this is not the case and DNA methylation is allele-specific. The first example is that in some cases DNA methylation is directly influenced by DNA sequence variation in *cis* and consequently shows Mendelian inheritance patterns (Schalkwyk, Meaburn et al. 2010). A second example, is seen in females where DNA methylation coordinates the random silencing of either the maternally- or paternally-derived X-chromosome to ensure dosage-compensation with males via the process of X-chromosome inactivation (XCI). A third type of allele-specific methylation is associated with regulating the monoallelic expression of imprinted genes. In most cases autosomal gene expression occurs from both alleles (i.e. paternal and maternal) equally, but at a small number of loci gene expression only occurs from one allele in a parental-origin-specific manner.

These genes are known as imprinted genes and occur in three variations. Some are expressed from only the maternally inherited allele and silenced on the paternally allele, some are only expressed from the paternally inherited allele and silenced on the maternal allele and some are isoform dependant whereby the allele which is expressed depends on the transcript of the gene. Imprinted genes are thought to have evolved with the placenta although parental effects on gene expression are also recognised to occur in insects (Crouse 1960) and plants (Kermicle 1970). The 'parental conflict' hypothesis, one of several theories to explain the evolution of genomic imprinting, postulates that imprinted loci evolved to regulate the competing interests of the parents while the developing offspring is *in-utero* (Moore and Haig 1991). The father's interest lies in providing the offspring with a high level of nutrients, thus ensuring survival and strong progeny. The mother's interests lie also in self-preservation and ensuring she has as much nutrients as she needs to survive also. Imprinted loci such as the *IGF2* are a good example of this struggle. *IGF2*, which is maternally imprinted and thus expressed form the paternal side, stimulates growth-promoting activity and larger offspring.

By silencing the maternal allele and providing mono-allelic expression of this gene, the mother can ensure the offspring receive an adequate nutrient intake without compromising her survival (Haig and Graham 1991; Wilkins and Haig 2003).

The number of imprinted genes of different status is shown in *Table 1.5*. There are ~65 confirmed imprinted genes in humans and ~75 in mouse, with considerable (but not complete) overlap (Jirtle 2006). The actual number of imprinted loci across the genome is likely to be higher; computational analyses suggest there may be over 600 (Luedi, Hartemink et al. 2005) and genome-wide studies of allele specific methylation suggest that additional imprinting DMRs may exist (Kerkel, Spadola et al. 2008; Schalkwyk, Meaburn et al. 2010). Furthermore, a recent study by Gregg *et al* uncovered evidence for widespread imprinting-like monoallelic expression across the brain (Gregg, Zhang et al. 2010), supporting the notion that these effects may be particularly important for neurobiological function (Wilkinson, Davies et al. 2007).

Table 1.5 - Imprinted Genes in Human and Mouse

The number of known classically-imprinted genes as well as predicted and unverified imprinted genes (Jirtle 2006)

Status	**Human**	**Mouse**
Conflicting Data	7	2
Imprinted	65	75
Not Imprinted	8	3
Predicted	111	2
Provisional Data	6	2
Unknown	15	9
Total	***212***	***93***

Many imprinted genes are situated near CpG islands, which are the target for the epigenetic modifications driving parental-origin-specific gene expression. These differentially methylated regions (DMRs) are methylated specifically on either the maternal or paternal

allele, and are often important for the control of imprinted genes. The grouping of imprinted genes within clusters allows them to share common regulatory elements such as DMRs, and when a DMR controls the monoallelic expression of one or more genes, it is known as an imprinting control region (ICR). This occurs in the case of *IGF2* and *H19* two reciprocally imprinted genes, which are regulated by the same ICR. On the maternal allele the ICR is unmethylated and allows the binding of CTCF. This blocks enhancers from binding to the promoter of *IGF2* and allows expression of *H19*. On the paternal allele, the ICR is methylated and so CTCF does not bind. This allows the enhancers downstream of *H19* to bind to the promoter of *IGF2* and enable expression (*Figure 1.8*) (Kurukuti, Tiwari et al. 2006). As will be discussed later in the chapter, altered patterns of DNA methylation across ICRs have been implicated in the aetiology of several neuropsychiatric diseases and may escape complete epigenetic reprogramming during gametogenesis (section 5.2.3).

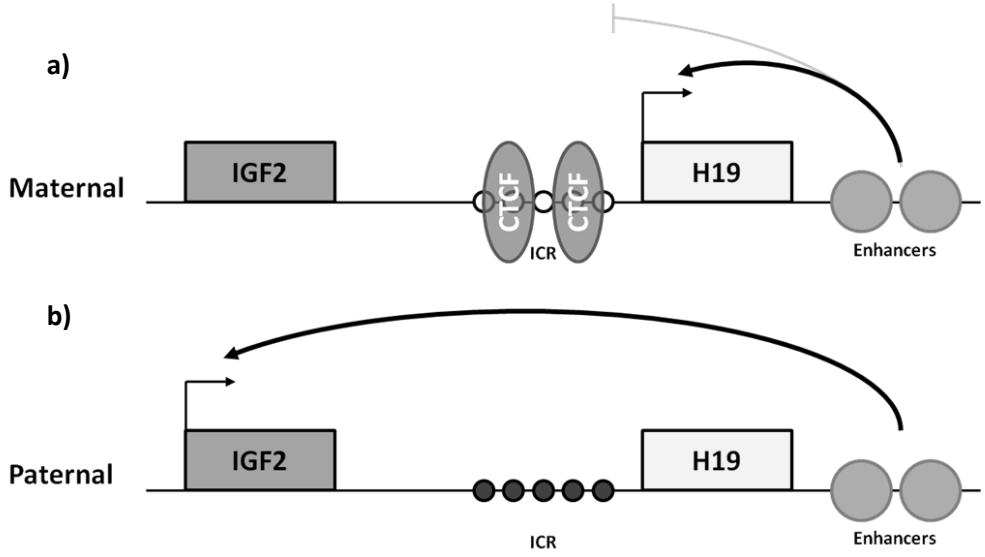


Figure 1.8 - IGF2/H19 ICR

The IGF2/H19 ICR has a downstream enhancer that activates either IGF2 or H19. In the maternally inherited allele, the ICR is unmethylated allowing for the binding of CTCF and blocks enhancers from accessing IGF2, causing the activation of H19 instead (a). In the paternal allele the ICR is methylated stopping the binding of CTCF and allows the enhancers to activate IGF2 instead of H19 (b).

Imprinted genes are involved in a variety of functions and many play a critical role in normal growth and development especially during gestation (Fowden, Sibley et al. 2006). Many classically-imprinted genes are expressed in the central nervous system and are known to play an important role in brain development (Wilkinson, Davies et al. 2007), behaviour (Tilghman 1999; Wilkinson, Davies et al. 2007; Curley, Mashoodh et al. 2010), and cognitive processes including language development and social function (Skuse, James et al. 1997). It has been demonstrated that loss of imprinting at the *IGF2* gene in the cerebellum is correlated with brain weight in males (Pidsley, Dempster et al. 2010), an interesting observation given the association between brain size and disorders such as schizophrenia (Harrison, Freemantle et al. 2003). Of note, several studies report that inter-individual differences in *IGF2* methylation may be environmentally mediated, particularly by prenatal exposure to famine (Heijmans, Tobi et al. 2008), which is of particular relevance given the known link between prenatal factors (including nutrition) and the risk of developing psychiatric illness as an adult. Any dysfunction in genetic imprinting could be extremely detrimental, causing over- or under-expression of a normally monoallelically-expressed gene or the expression of a gene from the wrong allele. Prader-Willi syndrome and Angelman syndrome, for example, both derive from the same deletion on chromosome region 15q11-13 (containing the paternally expressed genes *SNRPN* and *NDN*, and the maternally expressed gene *UBE3A*) but are characterised by very different symptoms. Prader-Willi syndrome, which is caused by a paternally-inherited deletion, was first described in 1956 and is characterised by obesity, dysmorphic appearance, short stature, underdeveloped genitals, developmental delays, mental retardation and behavioural problems (Cassidy 1997). Angelman syndrome, which is caused by a maternally inherited deletion, was first described in 1965 and is characterised by brachycephaly, dysmorphic appearance, hypopigmentation, jerky movement, delayed motor milestones, epilepsy and unusually happy disposition (Clayton-Smith and Pembrey 1992).

Of relevance to this project, low sperm counts have been associated with abnormal genetic imprinting and are more likely to transmit imprinting errors (Marques, Carvalho et al. 2004) or abnormal DNA methylation at imprinted sites (Kobayashi, Sato et al. 2007). As men of older age typically have lower sperm counts than men of a younger age (section 1.4.1), the potential for older fathers to pass on imprinting errors to their offspring may be higher. Several studies have shown that imprinted regions can harbour genetic variants that influence phenotype in a

parental-origin-specific manner, with opposite effects associated with maternally- and paternally-inherited alleles (Meaburn, Schalkwyk et al. 2010). For example, polymorphisms at the human *IGF2*/*H19* region have been shown to have a parent-of-origin specific association with placental and foetal growth (Adkins, Somes et al. 2010), and recent data shows that similar effects are observed for brain growth phenotypes and potentially neuropsychiatric disease (Pidsley, Dempster et al. 2012). Furthermore, parent-of-origin effects have been observed in autism (Cook, Lindgren et al. 1997; Schroer, Phelan et al. 1998), schizophrenia (Crow, DeLisi et al. 1989; Asherson, Walsh et al. 1994; Ohara, Xu et al. 1997) and bipolar disorder (McMahon, Stine et al. 1995). Many disease-associated risk-variants are located near known imprinted regions; in schizophrenia the 15q14-q14 region has been implicated (Freedman, Coon et al. 1997) which is in close proximity to the deletion seen in Prader-Willi/Angelman syndrome (Mann and Bartolomei 1999), and CNVs at the 15q locus have also been shown in autism (Cook, Lindgren et al. 1997; Schroer, Phelan et al. 1998). In Chapter 5, I will explore whether epigenetic changes at DMRs associated with several brain-expressed imprinted regions are associated with advanced paternal age.

1.6.3 Epigenetic Reprogramming and Transgenerational Epigenetic Inheritance

Mammalian germ line cells undergo extensive epigenetic reprogramming during development and gametogenesis to allow for zygotes to acquire totipotency of their cells (Sasaki and Matsui 2008). The overview of epigenetic programming in mouse is shown in *Figure 1.9*. Widespread erasure of DNA methylation occurs in early embryogenesis and DNA methylation marks are re-established in two phases. Once the genomes of the PGCs have been demethylated the cells enter mitotic (male) and meiotic (female) arrest.

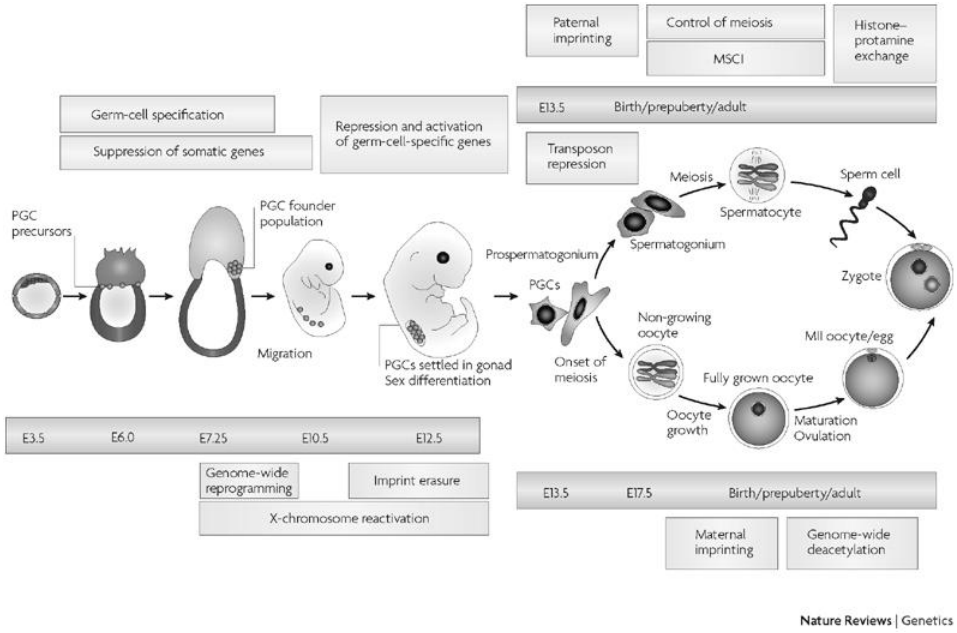


Figure 1.9 – Epigenetic Reprogramming in Mouse

Pattern of epigenetic reprogramming in germ cells during development of the gametes. Epigenetic reprogramming starts at E7.25 with the erasure of imprinting marks at E11. Imprinting is re-established at E13.5 in males and after birth in females. (Sasaki and Matsui 2008)

The first stage occurs before the development of the germ line cells and gives rise to the pattern of somatic cell like DNA hypermethylation in the cells of the pre-implantation embryo. Remethylation of the germ line cells occurs at different times in males and females. In males remethylation occurs at the prospermatogonia stage and precedes re-entry of the cells into mitosis (Kafri, Ariel et al. 1992). In females remethylation occurs during the growth of the oocytes (Kafri, Ariel et al. 1992). These then develop into all the somatic cells of the body. The second phase of the establishment of DNA methylation occurs in the primordial germ cells (PGCs). PGCs become demethylated in early development and the remethylation starts in prospermatagonia (*Figure 1.10*). This methylation pattern is hypomethylated compared to the somatic cells (Ariel, Cedar et al. 1994; Trasler 1998; Li 2002; Saitou, Barton et al. 2002; Santos and Dean 2004; Rousseaux, Caron et al. 2005; Biermann and Steger 2007; Oakes, La Salle et al. 2007). After fertilisation of the oocyte, the paternal genome is demethylated. Both the maternal and paternal genome is remethylated at implantation.

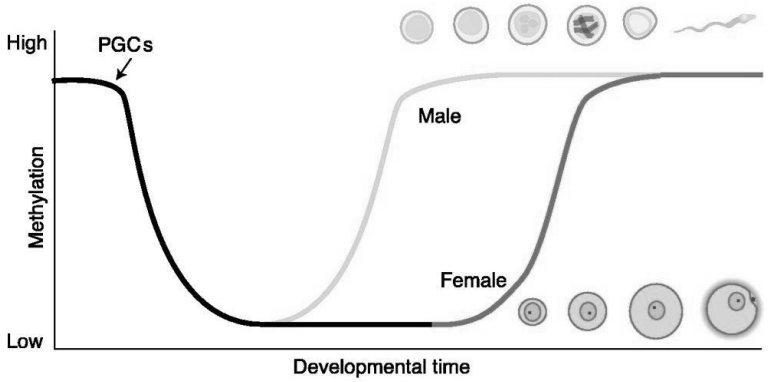


Figure 1.10 – Epigenetic Reprogramming in Germ Line Cells

Primordial Germ Cells become demethylated early in development. Remethylation begins in prospermatogonia in male germ cells and after birth in growing oocytes (Reik, Dean et al. 2001)

As summarised above, it is traditionally assumed that the epigenome is completely erased and re-established in the germ-line in order to allow totipotency in the developing embryo (Reik, Dean et al. 2001). Some recent studies, however, suggests that this resetting in primordial germ cells is not complete and there are elements which escape the total "wiping" of epigenetic information (*Figure 1.11*a). In particular, transposable elements such as IAP, long-interspersed nuclear element-1 (LINE-1), some imprinted loci and other elements in the genome (Figure 1.11b) have been shown to escape this resetting and retain some epigenetic modifications in the germ cells (Popp, Dean et al. 2010). These elements could provide a potential vehicle for transgenerational epigenetic inheritance. Although controversial, the notion that epigenetic marks may be heritable during meiosis in humans and thus potentially transmitted across generations (Klar 1998; Rakyan and Whitelaw 2003; Richards 2006) potentially blurs the demarcation between epigenetic- and DNA sequence–based inheritance, and challenges the assumption that the 'heritable' component to complex traits and disorders is entirely genetic. Therefore, if epigenetic differences arise as a function of advanced paternal age (either stochastically or environmentally), it is plausible that these changes could be transmitted to their offspring and potentially influence phenotype in the next generation. These ideas will be explored further in Chapter 5.

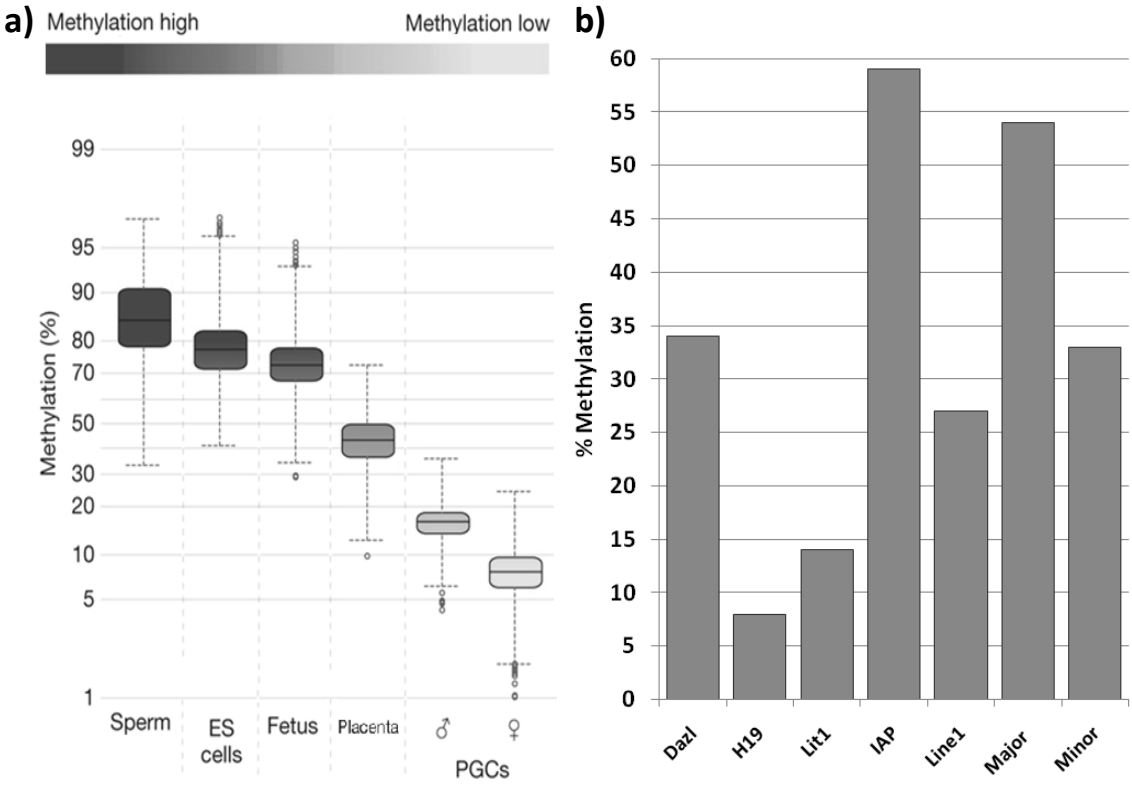


Figure 1.11 – Methylation Profile of Cells

Adapted from (Popp, Dean et al. 2010)

a) *Methylation profile in cells during development*
b) *Methylation of loci in E13.5 male primordial germ cells*
    i. *Dazl, H19, Lit1 (Kcnq1ot1) imprinted loci*
    ii. *IAP, Line-1 Transposable elements*
    iii. *Major and minor satellite repeats*

1.6.4 Environmentally- and Stochastically-Induced Epigenetic Variation

There is mounting evidence to suggest that epigenetic processes can be influenced by exposure to a range of environmental insults (Jirtle and Skinner 2007; Dolinoy and Jirtle 2008). DNA methylation, for example, has been shown to vary as a function of numerous nutritional, chemical, physical, and psychosocial factors. Imprinted genes have been shown to be particularly susceptible to epigenetic changes in response to the environment (Jirtle, Sander et al. 2000). Several studies report that inter-individual differences in methylation across DMRs regulating the monoallelic expression of imprinted domains may be environmentally

mediated. Individuals conceived during the Dutch Hunger Winter Famine of 1944–1945, for example, were found to be hypomethylated at the *IGF2* DMR on chromosome 11 (Heijmans, Kremer et al. 2007). In addition, random stochastic and developmental epigenetic changes are also likely to be important. Experiments tracking the inheritance of epigenetic marks through generations of genetically-identical cells in tissue-culture, for example, have indicated that there is considerable infidelity in the maintenance of methylation patterns in mammalian cells, and that *de novo* methylation events are fairly common during mitosis (Riggs, Xiong et al. 1998; Ushijima, Watanabe et al. 2003). As epigenetic processes are integral in determining when and where specific genes are expressed, such epigenetic metastability, environmentally- or stochastically-induced, may have profound phenotypic effects on gene expression in the cell.

These environmental and stochastic changes to the epigenome may explain the observation of epigenetic drift with age. There is a growing body of evidence suggesting that epigenetic variation in somatic cells is strongly associated with age (Flanagan, Popendikyte et al. 2006; Fraga and Esteller 2007; Siegmund, Connor et al. 2007; Bjornsson, Sigurdsson et al. 2008; Christensen, Houseman et al. 2009; Koch and Wagner 2011). At birth MZ twins have a very similar epigenetic profile, but as they age they exhibit increasing epigenetic differences (Fraga, Ballestar et al. 2005). Fraga and colleagues examined DNA methylation and histone acetylation in 80 pairs of MZ twins, ranging from 3 to 74 years of age, using a combination of global and locus-specific methods (Fraga, Ballestar et al. 2005). They found that one-third of MZ twins had a significantly dissimilar epigenetic profile, with older twins and those with a history of non-shared environments being the most disparate, suggesting that environmental factors may shape the epigenome over the life-course.

Of particular relevance to the study of advanced paternal age, sperm cells have also been shown to exhibit age-related DNA methylation changes (Flanagan, Popendikyte et al. 2006). Using a genome-wide approach to assess DNA methylation in sperm DNA from donors of a wide age range, 175 separate loci were identified that show age related changed in methylation; 88 showed a negative correlation with age, and 87 showed a positive association with age. The same study highlighted significant within-individual and between-individual epigenetic variability in male germline cells at CpG islands in the promoter regions of several

genes (Flanagan, Popendikyte et al. 2006). It is therefore reasonable to assume that with more cell divisions in older male germline cells, there is the potential for more environmentally- and/or stochastically-induced DNA methylation changes. Given the evidence discussed above regarding transgenerational epigenetic inheritance, it is plausible that some of these changes are passed on to subsequent generations. This hypothesis will be explored further in Chapter 5.

1.7 Current Animal Models of Parental Age

As previously discussed, animal models are powerful tools for investigating potential mechanisms involved in the development of disease. Currently there are only a small number of studies which look at the effects of paternal age in animal models. The first such study was by Auroux *et al* (Auroux 1983) who looked at behaviour in the offspring of rats of a variety of ages up to 23 months compared to the offspring of rats 2.5 months of age. They found that learning capacity, measured using an electrical stimulus learning task, was decreased in the offspring of older rats. The effect was continuous in male offspring, with the highest impairment being observed in the offspring of 22 month old fathers.

Garcia-Palomares *et al* (Garcia-Palomares, Pertusa et al. 2009) investigated paternal age in C57BL/6JIco mice, studying self-righting in postnatal day 4-6 pups, passive-avoidance and spontaneous motor activity. The offspring of old fathers showed decreased ability to self-right as pups, as well as reduced spontaneous motor activity in the home cage task and reduced passive-avoidance learning by using a light-dark box electric shock. A subsequent study by Foldi *et al* (Foldi, Eyles et al. 2010) in C56BL/6J mice showed that the female offspring of older fathers spent more time in the open arms of the elevated plus maze, increased exploration in the holeboard task and spent less time immobile in the forced-swim task.

From these studies, there is evidence for a change in behavioural phenotype associated with advanced paternal age. Only one of these studies looked at social behaviour but found no differences associated with paternal age (Foldi, Eyles et al. 2010) and so do not support the

human literature or the hypothesis that paternal age is associated with a social deficit (Weiser, Reichenberg et al. 2008).

1.8 Conclusions

From reviewing the epidemiological data looking into the role of parental age on the risk of developing psychiatric disease, it is clear that parental age effects are associated with the development of ASD, schizophrenia and BD and for all three disorders the evidence to support the paternal age effect is more consistent than with maternal age. Given the link between advancing paternal age and social deficit in adolescent males in the general population, and the common social deficits seen across ASD, schizophrenia and BD, it is plausible that paternal age may not be necessarily directly related to each disorder but rather mediate the social deficit phenotype shared across them all.

The potential mechanism(s) underlying the paternal age effect are not known. Two plausible mechanisms include an increased burden of *de novo* CNVs or epigenetic changes occurring in the male germline. Recurrent *de novo* CNVs have been shown to cause a large percentage of cases of ASD, schizophrenia and BD and the higher rate of CNVs is expected with advanced paternal age due to a higher mutation rate and increased DNA fragmentation in sperm of older males. Although no conclusive association has been reported between paternal age and CNV load, this is a potential cause of the paternal age effect that warrants further investigation.

Epigenetic dysfunction is another plausible mechanistic route between advanced paternal age and risk for neuropsychiatric disease, in particular with regard to imprinting dysfunction and DNA methylation across transposable elements which are known to be labile in relation to age and potentially escape total epigenomic reprogramming during gametogenesis. For example, because the sperm of older fathers have more time to be exposed to environmental insults and accrue stochastic alterations over multiple cell divisions before fertilisation, the potential for age-related changes in germ-line DNA methylation is increased.

1.9 Aims of the Thesis

In this thesis I hope to establish an animal model of advanced paternal age to investigate whether any specific behavioural effects are observed in the offspring of older fathers. In particular, I am interested to explore where any observed behavioural changes are specific to social phenotypes or whether advanced paternal age has a more general effect on behaviour. I will then use this model to explore the potential mechanism(s) by which the paternal age effect is mediated. This includes investigating whether advanced paternal age causes a greater load of *de novo* CNVs or a higher rate of recurrent CNVs using a genome-wide assessment of genomic variation. I also aim to explore epigenetic changes associated with advanced paternal age, by examining global changes in DNA methylation, and also focussing on DMRs regulating the transcription of brain expressed imprinted genes and transposable elements such as IAP and LINE-1 elements. Finally, using genome-wide gene expression analysis, I aim to uncover transcriptomic differences between the offspring of older and young fathers to potentially identify candidate genes for the cause of the paternal age effect and the disorders it affects.

Chapter 2 - General Materials & Methods

2.1 Introduction

This section will provide details about general methods that are used across multiple chapters throughout my thesis and will be referred back to where relevant. Additional methods specific to individual chapters will be described throughout the thesis.

2.2 Extraction of Nucleic Acids

DNA and RNA was isolated from a variety of tissue sources including spleen, brain and sperm, and used in a range of methodologies throughout the research presented in this thesis. The general principal of nucleic acid purification is to remove all proteins and other impurities to leave only pure high molecular weight DNA or RNA that can be used for downstream processes. Depending upon the method of extraction and the tissue used, up to 30ug of DNA can be obtained from each 10mg of animal tissue. A variety of methods can be used for isolating DNA and RNA, and below I outline the methods used in this thesis.

Table 2.1 – Methods used for Nucleic Acid Extractions

Tissue	Nucleic Acid	Extraction Method	Experiment
Spleen	DNA	Qiagen DNeasy (Section 2.2.1)	CNV (Chapter 4) Epigenetics (Chapter 5)
Frontal Cortex	DNA RNA	Qiagen AllPrep RNA/DNA Extractions (Section 2.2.2)	Epigenetics (Chapter 5) Expression (Chapter 6)
Cerebellum	DNA	Phenol-Chloroform (Section 2.2.3)	Epigenetics (Chapter 5)
Hippocampus	DNA	Phenol-Chloroform (Section 2.2.3)	Epigenetics (Chapter 5)

2.2.1 Qiagen DNeasy Columns

All extractions were carried out on between 20mg of frozen tissue and using a protocol adapted from the standard Qiagen DNeasy protocol (Qiagen 2011). Before starting, buffers

were prepared by adding 25ml 100% ethanol to AW1 and 30ml 100% ethanol to AW2. Using a clean scalpel 20mg of tissue was excised from main sample and put in a clean eppendorf tube. The excised tissue was then homogenised with a DNase free, sterile, plastic pestle. 180µl of ALT buffer and 20µl Proteinase K (PK) (20mg/ml) was added to the homogenised sample and drawn up into a pipette tip to wash any tissue remaining from the pestle, the sample was then mixed by vortexing. The homogenised sample with PK was incubated at 56^{o}C in a water bath for 2-3hrs with vortexing every 30mins until tissue was completely lysed. After the tissue was completely lysed, it was vortexed for 15secs. 200µl of buffer AL was then added to the lysed tissue and mixed by vortexing followed by the addition of 200µl of 100% ethanol, the sample was then mixed again by vortexing.

The sample was then pipetted on to the DNeasy mini spin column sat on a 2ml collection tube. The column was spun at 8000rpm for 1min. The collection tube was discarded and put column on new collection tube to which 500µl of buffer AW1 was added and the column was then spun again at 8000rpm for 1min. Again the collection tube was discarded and the column put on new collection tube. 500µl of buffer AW2 was added to the column and spun at 8000rpm for 3mins. The final collection tube was discarded and the column put on to a clean eppendorf tube. 50µl of DNase free water was applied directly to the membrane of the spin column and incubated for 3 minutes at room temperature after which it was spun at 8000rpm for 1min. Another 50µl of DNase free water was applied directly to the membrane of the spin column and incubated for a further 3 minutes and spun at 8000rpm for 1min. Samples were then quantified using NanoDrop NO-1000 using the DNA-50 analysis type checking 260/280 ratio was over 1.8 and 260/230 ratio was around 2. The quality and integrity of DNA was also checked using a 0.8% agarose gel using gel electrophoresis (*Figure 2.1*) (section 2.5).

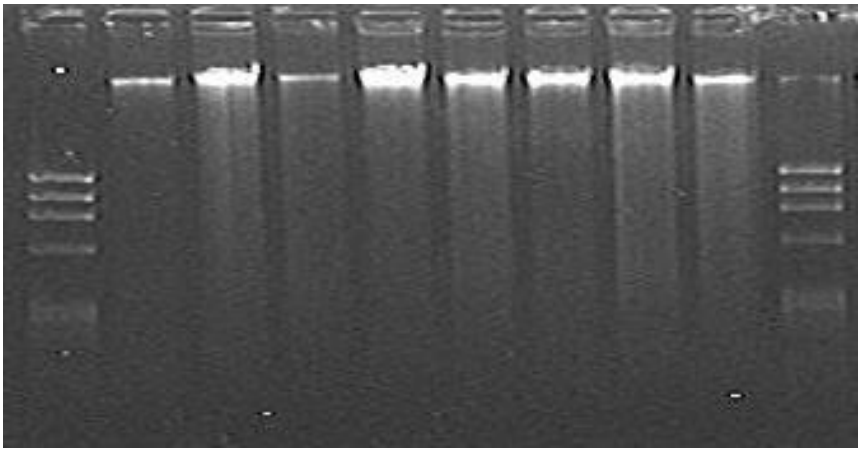


Figure 2.1 - Example of spleen DNA extracted using Qiagen DNeasy protocol

Ladder used is ΦX174 DNA-HaeIII ladder

2.2.2 Qiagen AllPrep RNA/DNA Extractions

DNA extractions for the frontal cortex were carried out using the Qiagen AllPrep RNA/DNA kit using a protocol modified from the manufacturer's standard method to maximise yield from brain tissue. 30mg of frontal cortical tissue was homogenized on dry ice with a DNase free, sterile, plastic pestle. 350µl of RLT buffer was added after which the lysate was passed through a 20-gauge needle 10 times. The lysate was centrifuged for 3 minutes at 13000rpm. The supernatant was carefully pipetted off and transferred to an AllPrep DNA spin column placed in a 2ml collection tube. The spin column and tube were then centrifuged for 30 seconds at 13000rpm. The flow through was then pipetted back on to the DNA spin column and centrifuged for a further 30 seconds at 13000rpm. The AllPrep DNA spin column was placed on to a new collection tube, and stored on ice until later DNA purification steps.

350µl of 70% ethanol was added to the flow through from the DNA spin column and mixed well by pipetting. This was then transferred to an RNeasy spin column placed in a 2ml collection tube and centrifuged for 15 seconds at 13000rpm. The flow-through was discarded and 700µl of Buffer RW1 was added to the RNeasy spin column then centrifuged 15 seconds at 13000rpm. The flow through was again discarded and 500µl Buffer RPE was added to the RNeasy spin column and centrifuged for 15 seconds at 13000rpm. The flow through was discarded and another 500µl of Buffer RPE was added to the RNeasy spin column and

centrifuged for 2 minutes at 13000rpm. The RNeasy spin column was placed on a new 2ml collection tube and centrifuged at 13000rpm for 1 minute to fully remove any residual liquid. The RNeasy spin column was placed on a new 1.5ml collection tube and 30µl of RNase-free water was directly added to the spin column membrane and centrifuged for 1 minute at 13000rpm to elute the RNA. Another 30µl of RNase-free water was directly added to the spin column membrane and centrifuged for 1 minute at 13000rpm. For long term storage 75µl of 100% ethanol was added to the RNA, and the precipitated sample stored at -80^{o}C.

500µl Buffer AW1 was added to the AllPrep DNA spin column and centrifuged for 15 seconds at 13000 rpm. The flow through was discarded and 500µl Buffer AW2 was added to the AllPrep DNA spin column and centrifuged for 2 minutes at 13000rpm. The AllPrep DNA spin column was placed on a new 1.5ml collection tube and 100µl RNase-free water was added directly to the spin column membrane and incubated at room temperature for 5 minutes. This was then centrifuged for 1 minute at 13000rpm to elute the DNA. Another 50µl of RNase-free water was added directly to the spin column membrane and incubated at room temperature for 5 minutes. This was then centrifuged for 1 minute at 13000rpm to elute the DNA.

DNA and RNA samples were quantified using a Nanodrop spectrophotometer and checked for an OD230nm/OD280nm of over 1.8. The DNA and RNA were then checked on a 0.8% agarose gel for size and fragmentation (*Figure 2.2*) (section 2.5). RNA was also assessed using the Agilent Bioanalyzer system (see section 6.4.1) to ensure high-quality samples for microarray analysis.

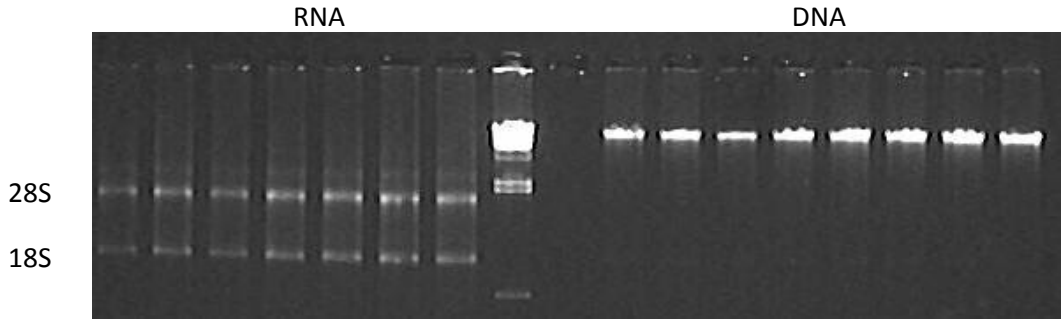


Figure 2.2 - RNA and DNA Extractions on 0.8% Agarose Gel

2.2.3 DNA Extraction using Phenol-Chloroform

All extractions were carried out on between 10mg and 50mg of frozen tissue. Before starting, lysis buffer (10mM Tris-HCl, 0.1M EDTA, 0.5% SDS) was made and autoclaved. Using a clean scalpel, tissue was excised from the main sample and put in a clean eppendorf tube. The excised tissue was then homogenised using a DNase free, sterile plastic pestle. 300ml of lysis buffer and 3.3ul PK (18.5mg/ml) were then added to the homogenised tissue and the samples were incubated for 16 hours in a water bath set at 50^{o}C. After the incubation, samples were spun down and pipetted up and down with a p1000 pipette to break up any unlysed tissue. Another 1.65ul PK (18.5mg/ml) was added to the sample followed by another hour at 50^{o}C in the water bath. To inactivate the PK, samples were incubated in the water bath at 65^{o}C for 30 minutes after which samples were cooled to room temperature.

Lysed samples were transferred to phase lock tubes made by adding vacuum grease to a 2ml eppendorf tube and spun down. 300ml of phenol-chloroform-isoamyl alcohol (PCI) buffered to pH 7.5 was added to the phase lock tube and the samples were inverted 20 times then spun in a centrifuge at 13000rpm for 15 minutes. The aqueous layer was carefully pipetted off the top of the phase lock layer and put in clean eppendorf tube. 300ml of chloroform was then added to the removed aqueous layer and inverted 20 times after which the sample was spun in the centrifuge at 13000rpm for 15 minutes. The aqueous layer was carefully pipetted off the interphase layer and put into a clean eppendorf tube and another 300ml of chloroform was added and inverted 20 times. The sample was spun again in centrifuge at 13000rpm for 15 minutes. The aqueous layer was carefully pipetted off and put in a clean eppendorf tube. 75ul of 10M Ammonium acetate (NH_4Ac) and 600ml ice cold 100% ethanol were added to each sample which are inverted slowly 20 times and then left at -20^{o}C in the freezer for 1 hour.

After the incubation, samples were spun in a centrifuge at 13000rpm for 15 minutes. The supernatant was removed and 1ml of 70% ethanol was added then the sample was spun in a centrifuge at 13000rpm for 10 minutes. The supernatant was again removed and eppendorfs were inverted open on tissue allowing the pellet to air dry for 1 hour. Samples were resuspended in 100ul of DNase free water. Samples were then quantified using NanoDrop NO-1000 using the DNA-50 analysis type checking 260/280 ratio was over 1.8 and 260/230 ratio

was around 2. Quality and size of DNA was checked using a 0.8% agarose gel using gel electrophoresis (section 2.5).

2.3 Sodium Bisulfite Conversion

The 'gold standard' method for mapping methylated cytosines involves the treatment of genomic DNA with sodium bisulfite; this process converts unmethylated cytosines to uracils (and subsequently, via PCR, to thymidines), while methylated cytosines are resistant to bisulfite and remain unchanged (Clark, Harrison et al. 1994) (*Figure 2.3*). After sodium bisulfite treatment, DNA regions of interest are amplified and interrogated to identify C → T transitions or stable C positions, respectively corresponding to unmethylated and methylated cytosines in the native DNA. Numerous methods of analyzing bisulfite-modified DNA have been described (see section 5.4.6).

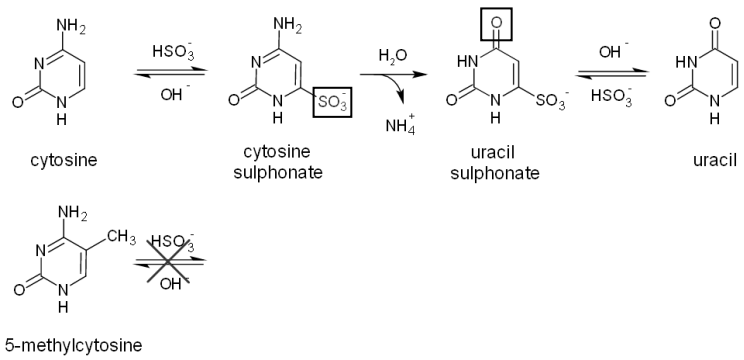


Figure 2.3 - Bisulfite Conversion of Unmethylated Cytosine Residues

(Wikipedia 2011)

Sodium bisulfite conversion was performed using the EZ-96 DNA Methylation Kit from Zymo Research (ZymoResearch 2011). First, the CT Conversion Reagent was prepared by adding 7.5ml of DNase free water and 2.1ml of M-Dilution Buffer which was then mixed at room temperature by vortexing for 10 minutes. M-Wash Buffer was prepared by adding 144ml of 100% ethanol to the 36ml of concentrate. 500ng of high molecular weight DNA was diluted in 45µl DNase free water (11.1ng/µl) in a conversion plate. 5µl of the prepared M-Dilution Buffer was added to each well of the conversion plate and mixed by pipetting. The conversion plate

containing the samples was incubated on a thermocycler at 37^{o}C for 15 minutes. After the incubation, 100µl of the prepared CT Conversion Reagent was added to each sample and mixed by pipetting. The conversion plate was then incubated in the dark at 50^{o}C for 16 hours using a thermocycler after which the plate was incubated at 2^{o}C for 10 minutes.

400µl of M-Binding Buffer was added to each well of the Silicon-A Binding Plate on a Collection Plate then the samples from the conversion plate were added into the wells of the Silicon-A Binding Plate containing the M-Binding Buffer. This was mixed by pipetting then centrifuged at 4000 x g for 5 minutes after which the flow through was discarded. 500µl of M-Wash Buffer was added to each well and centrifuged at 4000 x g for 5 minutes. 200µl of M-Desulphonation Buffer was added to each well and incubated at room temperature for 20 minutes after which the plate was centrifuged at 4000 x g for 5 minutes. 500µl of M-Wash Buffer was added to each well and centrifuged at 4000 x g for 5 minutes. Another 500µl of M-Wash Buffer was added to each well and centrifuged for 10 minutes 4000 x g to fully dry the membranes of the Silicon-A Binding Plate.

The Silicon-A Binding Plate was placed on to an elution plate and 30ul of M-Elution Buffer was applied directly to the binding matrices of the Silicon-A Binding Plate. It was then centrifuged for 3 minutes at 4000 x g to elute the DNA. Another 30ul of M-Elution Buffer was added directly to the binding matrix in each well of the Silicon-A Binding Plate and centrifuged for 3 minutes at 4000 x g to elute the DNA. Samples were then quantified using NanoDrop NO-1000 using the ssDNA-33 for single stranded analysis type checking 260/280 ratio was over 1.8 and 260/230 ratio was around 2. DNA was then plated out for polymerase chain reaction (PCR) using 1µl of converted DNA and stored at -20^{o}C. For longer term storage, plate was stored at -80^{o}C as bisulfite treated DNA is less stable due to being single stranded DNA (ssDNA).

2.4 Polymerase Chain Reaction (PCR)

PCR is used to amplify short sections of DNA to produce large amounts of a target-specific sequence using short priming sequences of ssDNA. The main elements of PCR and their

function are outlined in *Table 2.2*. For each specific PCR, different annealing temperatures are required as well as different amounts of magnesium chloride ($MgCl_2$), which is determined by $MgCl_2$ titration. Below is listed a standard PCR recipe on which all PCRs in my thesis are based on. Assay-specific optimised protocols (where relevant) are given in the relevant sections of the thesis. Qiagen Hot-Star Taq was used in bisulfite PCRs as hot-start PCR thermocycling conditions are optimal for bisulfite PCR reactions.

PCR Mix	*1x*
10 X Reaction Buffer	1µl
Magnesium Chloride (25 mM)	0.2µl
Deoxynucleoside triphosphates (dNTPs) (2.5mM)	0.2µl
Qiagen Hot Star *Taq* Polymerase (5 units/µl)	0.05µl
Primer Mix (10uM)	1µl
Water	6.55µl
DNA (10ng/µl)	1µl

Table 2.2 - Table of PCR Components and their Function

Component	**Function**
Genomic DNA	Provides the template for amplification. In the case of bisulfite-PCR amplification, bisulfite-converted genomic DNA is used as the reaction template
PCR Buffer	Contains agents to maintain an optimum pH for amplification
$MgCl_2$	Provides required divalent cations and stabilizes DNA strands. Lower concentrations of $MgCl_2$ also lead to more specific PCRs
Primers	Small sequences of ssDNA complementary to the region of interest which anneal to the DNA to provide an initiation site for PCR
Taq Polymerase	Enzyme which allows for subsequent amplifications of DNA due to thermostability. Produces complementary strand to template strand beyond the primers
dNTPS	Free individual nucleotide bases which are used to form the complementary strand by the Taq polymerase

After the PCR mix was made and plated out, the samples were then transferred to a thermocycler and the below PCR cycling program was used. The annealing temperature defines the temperature at which the primers anneal to the DNA template strands. The initial

95°C incubation is required to activate the Hot Star Taq polymerase. The second 95°C at the start of each cycle is required to denature the DNA strands into ssDNA. Cooling to the annealing temperature causes the primers to anneal to the relevant region of DNA to be amplified and is sequence-specific, for example varying as a function of the GC content of the primer sequences. Annealing temperatures are critical for optimal PCR amplification: if they are set too high, the primers won't anneal to the DNA and no amplification will occur. If they are too low, the primers will anneal to DNA with low specificity and will result in non-specific amplification. The 72°C at the end of each cycle instigates the extension step and allows for the Taq polymerase to build the complementary strand beyond the primers using the dNTPs.

Cycling Conditions

95°C	15 minutes	
95°C	20 seconds	←
Annealing Temperature	20 seconds	35 times
72°C	30 seconds	←
72°C	3 minutes	
4°C	10 minutes	

All PCR primers were designed for an annealing temperature of 55°C but some PCRs required optimizing by adjusting the annealing temperature or the amount of $MgCl_2$ added due to weak or no PCR product or non-specific amplification. Again, assay-specific conditions are given in the relevant sections of the thesis.

2.5 Agarose Gel Electrophoresis

Agarose gel electrophoresis provides a quick method for the analysis of DNA quality and to assess PCR products based on their size. By applying an electrical charge across the gel, negatively charged DNA moves towards the positive terminal. Agarose powder is melted in 1 x Tris-borate-EDTA (TBE) buffer and the concentration of agarose is used to control the speed at which DNA moves through the gel. Large DNA molecules (such as those expected after successful DNA extraction) migrate relatively slowly and so are generally checked on a 0.8-1.0% agarose gel. Small DNA fragments (such as after sonication or shearing) normally move

through gels quickly and so are checked on a higher concentration gel of ~3%. The PCR products assessed in this thesis were generally checked on a 1.5-2% gel. DNA bands were visualised by the addition of ethidium bromide, which integrates into double-stranded DNA and is visible under UV light. All DNA was run against a ΦX174 DNA-HaeIII ladder unless otherwise stated and loaded with Orange G loading dye.

Chapter 3 - Behavioural Assessment in a Mouse Model of Advanced Paternal Age

3.1 Abstract

Previous epidemiological studies have reported paternal age to be a risk factor for several human neuropsychiatric disorders including ASD, schizophrenia and BD. A common phenotype across all these disorders is social dysfunction. It is not known if other factors such as maternal age, other environmental factors such as prematurity or genetics contribute to the manifestation of this phenotype. To control for genetics, environment and maternal age, and to investigate any behavioural changes associated with paternal age, male C57BL/6J mice of three different ages (2 months, 10 months and 12 months) were bred with 2 month old C57BL/6J females and the resulting offspring were tested for locomotor activity, anxiety, exploration, short and long term memory and social interaction in a battery of behavioural tasks. It was observed that the offspring of old (10 month old) fathers engaged in significantly less exploratory and social behaviour than the offspring of standard breeding age males (2 months old). These observations provide evidence that advanced paternal age in mouse is a contributing factor to behavioural changes without variations in maternal age, living environment or genetics.

3.2 Introduction

As discussed in detail in the introductory chapter (section 1.2.2), advanced paternal age has previously been reported as a risk factor for ASD, schizophrenia and BD. ASDs are characterised by multiple symptoms but primarily by impairments or delays in social functioning, as first described by Kanner in 1968 (Kanner 1968). Deficits in social behaviour are also a characteristic of other psychiatric disorders. Social impairment and being socially withdrawn are clinical symptoms of schizophrenia (Mueser, Bellack et al. 1991) and shyness or social anxieties are characteristics of BD (Simon, Otto et al. 2004). As previously discussed, however, it is not known whether paternal age is directly associated with ASD, schizophrenia and BD, or whether it is a risk factor for the common phenotype of social dysfunction that transcends all these disorders.

As the association between paternal age and psychiatric conditions has only been observed in epidemiological human studies, there are many limitations. First, a limitation in all human based studies is the potential confounding effect of genetic variation. There are many

studies into the genetic causes of ASD, schizophrenia and BD but separating any inherited genetic effects from purely paternal age effects is very difficult, especially given the link between *de novo* mutations and paternal age (section 1.5). Although many candidate genes and genomic regions have been identified for ASD (Vorstman, Staal et al. 2006; Ingason, Kirov et al. 2011), schizophrenia (Stefansson, Rujescu et al. 2008; Stone, O'Donovan et al. 2008) and BD (Serretti and Mandelli 2008; Zhang, Cheng et al. 2009) (section 1.5), genetic associations are characterised by considerable heterogeneity and small effect sizes, and none has been shown to be directly causal for any of these complex disorders. Furthermore, genetic variation in the general population could possibly confound any associations with paternal age via mechanisms such as population stratification. As genetic variation cannot be easily controlled for in human studies, it cannot be determined if any/all of the observed paternal age effects observed are confounded by genetic variation.

Second, the effect of the environment also needs to be taken into account. For example, foetal distress (Glasson, Bower et al. 2004), smoking during pregnancy (Hultman, Sparen et al. 2002), preterm birth and low birth weight (Kolevzon, Gross et al. 2007) have all been suggested as risk factors for the development of ASD. In schizophrenia, social-economic status (Werner, Malaspina et al. 2007), prenatal malnutrition (Susser, Neugebauer et al. 1996) and early exposure to cannabis (Andreasson, Allebeck et al. 1987) have all been reported as risk factors and in BD, stressful life events are a risk factor (Alloy, Abramson et al. 2005). In human epidemiological studies it is extremely difficult to control for the effect of specific environmental exposures, whether they have been previously considered a risk factor or not. Again, as all environmental exposures cannot be controlled for, it is not entirely clear if paternal age effects could be the result of correlated environmental influences rather than, or in addition to, genetic variation.

Third, many paternal age effects are reported in conjunction with maternal age effects and so it is not clear whether maternal age is the true risk factor for the development of these disorders instead of paternal age (Lopez-Castroman, Gomez et al. 2010) or whether it is the combination of both advanced maternal and paternal ages (Gillberg 1980; Croen, Najjar et al. 2007; Durkin, Maenner et al. 2008; Grether, Anderson et al. 2009). Although late fatherhood often coincides with advanced maternal age, this doesn't explain the more consistent

epidemiological findings seen with paternal age than with maternal age (Lauritsen, Pedersen et al. 2005; Reichenberg, Gross et al. 2006). Furthermore, advanced paternal age is likely to be more of a widespread risk factor in the general population due to extended male fertility compared to women.

For the reasons discussed, the development of an animal model would be extremely useful in determining the true role of advanced paternal age in mediating behavioural changes associated with psychiatric disorders. In animal models genetic variation can be controlled by using inbred strains established over multiple generations to create animals which are homozygous at all loci except the sex chromosomes, minimising the amount of genetic variation due to SNPs and VNTRs. The use of animal models also means that the environment can be tightly controlled and any large variations in diet, temperature, humidity, light exposure and stressful events can be avoided. Any environmental events which could potentially cause changes in behaviour can be recorded and taken into account when analysing behavioural results. Finally, the effect of maternal age can be directly controlled for and the outcome of only paternal age observed by breeding males of various ages with a consistent age of female. In using animal models, many of the limitations of human epidemiological studies can thus be overcome.

As well as overcoming epidemiological limitations, the utility of animal models has many other benefits for this study. As mice are small mammals, the breeding, gestation and development of test animals can be carried out quickly whilst still controlling for previously discussed issues. Typical breeding age for mice starts around two months and male breeders are generally retired after 7–8 months when they are considered past their reproductive peak, so advanced paternal age does not have to be defined by large intervals in age. Animals are also useful because measures of specific behaviours can be selected that are of interest for the disorders being studied. For example, because a common phenotype in ASD, schizophrenia and BD relates to deficits in social interaction, social behavioural tasks were selected for this study. Other behaviours can also be tested to see if there are more general effects of paternal age, or whether these are limited to specific behaviours.

Autistic-like reductions in social behaviour have been previously modelled in rodents. In their study, Moy *et al* showed that certain strains of mouse (such as DBA/2J and A/J) show less preference for social interactions and engage in less overall social activity (Moy, Nadler et al. 2004). Previous animal studies of paternal age have found that advanced paternal age has a detrimental effect on behaviour although only one looked at social behaviour (Foldi, Eyles et al. 2010) and found that it is affected. Some behavioural changes previously shown to associate with advanced paternal age are reduced learning capacity (Auroux 1983), decreased ability to self right on postnatal day 4-6, reduced spontaneous motor activity and reduced passive-avoidance learning (Garcia-Palomares, Pertusa et al. 2009). Other behaviours shown to be affected by advanced paternal age in animal models which are not detrimental are increased exploration in both holeboard and the elevated plus maze (EPM) (Foldi, Eyles et al. 2010).

3.3 Aims

The overall goal of this chapter is to investigate whether advanced paternal age is associated with behavioural changes in the offspring using a rodent model that allows us to control for parental genetic variation, fluctuations in maternal age or changes to environment. A secondary aim was to explore whether any behavioural changes observed are limited to those related to ASD, schizophrenia and BD or whether there are more global effects on behaviour.

3.4 Methods

3.4.1 Breeding Strategy

All animals were bred and maintained in the Biological Services Unit at the Institute of Psychiatry, Kings College London using stocks purchased from Charles River Laboratories (Margate, U.K.). All housing and experimental procedures were performed in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986. C57BL/6J inbred mice were used in this study as it is the most commonly used strain of laboratory mouse and was the first strain to have its genome completely sequenced. Therefore, the genome of this strain is also well characterised and annotated in genome browsers. It has been used in many previous

behavioural studies and performs well in the tasks chosen including the social behaviour and discrimination tests (Moy, Nadler et al. 2004).

For breeding, females aged two months were primarily bred with males of two different ages; 'young males' of normal breeding age of two months and 'old males' (retired breeders), 10 months old. Furthermore, females were bred with a smaller number of older retired breeders or 'very old males', 12 months in age (*Figure 3.1*). 14 out of 15 breeding pairs produced usable litters where more than one male offspring was born, four litters for young male breeders (and one producing no offspring), six litters for old male breeders and four litters for very old male breeders. The average litter size within the young and old male groups was seven (male to female ratio 1:1) and in the very old male group the average litter size was eight. The total progeny generated was 40 mice in the young fathers group, 44 mice in the old fathers group and 31 in the very old male group. Of these there were 19 males in the young father group, of which one had to be culled due to abnormal development, 24 in the old father group and 12 males in the very old father group, of which one had to be culled due to hydrocephaly (*Table 3.1*). Only male offspring were kept due to the sex ratio in autism being 4:1 (Fombonne 2003). All female offspring were culled at weaning leaving only male offspring for testing.

Figure 3.1 - Diagram of Breeding Strategy

Young male n=5, Old male n=6, very old male n=4, female n=15

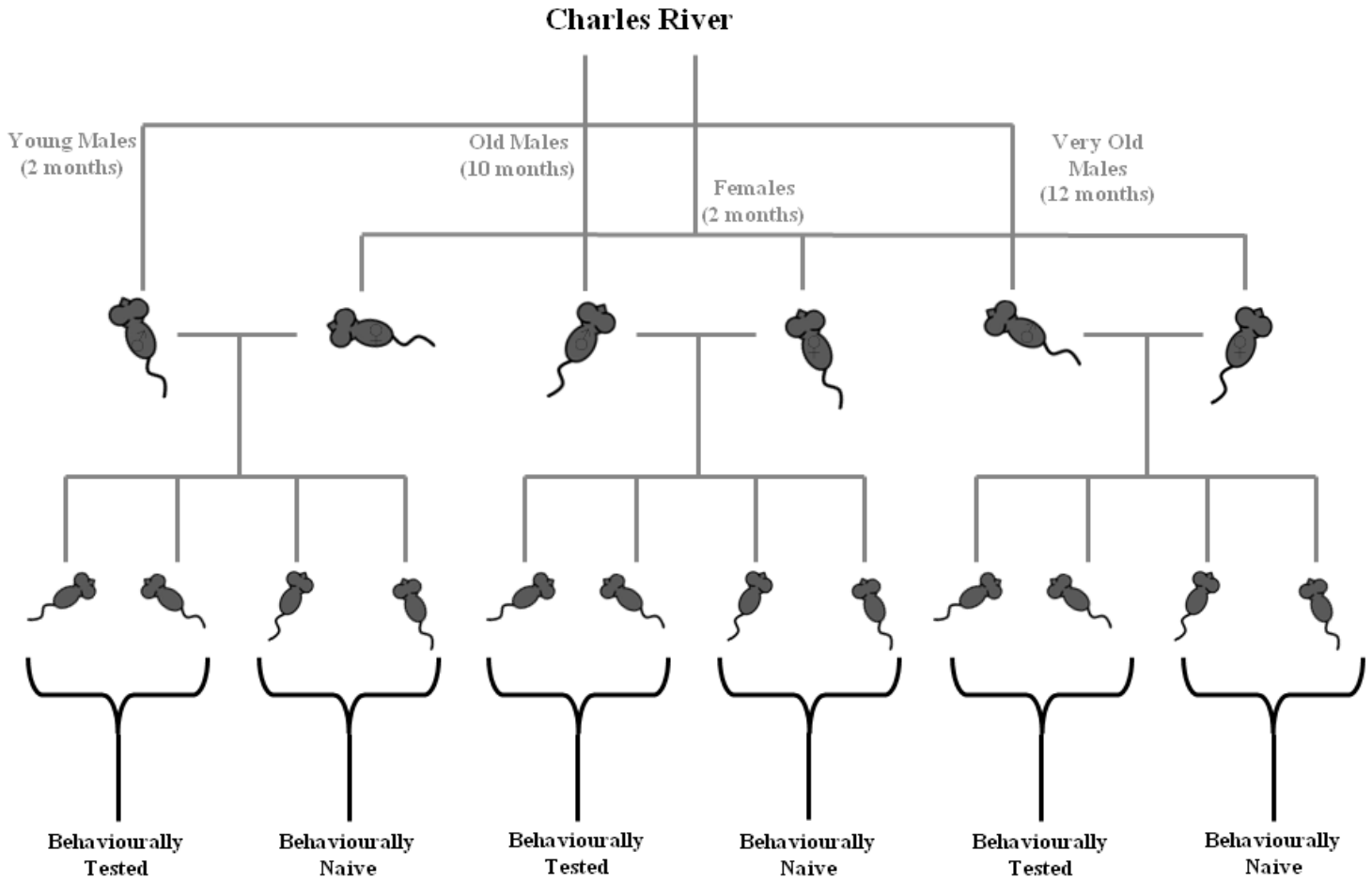


Table 3.1 – Individual Litter Details

Pair ID	Group	Breeder ID	Sex	Age (months)	Litter born	Litter size	M	F	Comments
1	Young Male	21	F	2	02/12/08	7	4	3	
		11	M	2					
3	Young Male	27	F	2	03/12/08	9	2	7	
		13	M	2					
4	Young Male	30	F	2	01/12/08	10	6	4	x2 cages of 3 offspring
		14	M	2					
5	Young Male	33	F	2	03/12/08	0	0	0	
		15	M	2					
6	Young Male	36	F	2	02/12/08	8	7	1	x2 cages of 3 offspring
		16	M	2					1 male runt culled
8	Very Old Male	22	F	2	02/12/08	7	3	4	1 male culled (hydrocephaly)
		2B	M	12					
10	Very Old Male	28	F	2	01/12/08	8	3	5	
		6B	M	12					
13	Very Old Male	37	F	2	03/12/08	8	3	5	
		12B	M	12					

Table 3.1 Continued

Pair ID	Group	Breeder ID	Sex	Age (months)	Litter born	Litter size	M	F	Comments
14	Very Old Male	40	F	2	01/12/08	8	3	5	
		14B	M	12					
15	Old Male	23	F	2	01/12/08	6	3	3	
		1	M	10					
16	Old Male	26	F	2	01/12/08	7	4	3	
		2	M	10					
17	Old Male	29	F	2	01/12/08	7	4	3	
		3	M	10					
18	Old Male	32	F	2	02/12/08	10	5	5	
		4	M	10					
19	Old Male	35	F	2	01/12/08	7	5	2	
		5	M	10					
20	Old Male	38	F	2	01/12/08	7	3	4	
		6	M	10					

The offspring were weaned aged 4–5 weeks and group housed with their siblings unless males in the litter exceeded 4, in which case the litter was split in two. Mice were individually housed for two weeks prior to testing. Mice were housed in standard cages measuring 30.5×13×11 cm, with food and water available *ad libitum*. The housing room was maintained on a standard light/dark cycle with white lights on from 08:00 to 20:00. Ambient temperature in all rooms was maintained at 21±2°C with 45% humidity.

3.4.2 Offspring Behavioural Testing

Two to three males were randomly selected from each litter for behavioural testing giving 12 offspring from the young and old father breeding groups and nine from the very old father breeder group. Mice were randomised within their groups for testing. Offspring were aged 12 weeks at the start of testing and all testing took place during the light phase with a light level less than 30 lux in the test room. Each apparatus was wiped clean with 1% Trigene® between subjects to avoid olfactory cueing behaviours. Behaviours for all tests were recorded on videotapes for further detailed analysis. Mice were returned to their home cage at the end of each test. The tests were carried out as per the schedule in *Table 3.2*.

Table 3.2 - Behavioural Testing Diary

Month	Date	Test	Trial	Batch
February	23	Open Field		Batch 1
	24	Novel Object	0hr and 1hr trial	Batch 1
	25	Novel Object	24hr trial	Batch 1
	26			
	27	Holeboard		Batch 1
	28			
March	1			
	2	Open Field		Batch 2
	3	Novel Object	0hr and 1hr trial	Batch 2
	4	Novel Object	24hr trial	Batch 2
	5			
	6	Holeboard		Batch 2
	7 8			
	23	Social Investigation and Discrimination	Habituation	Batch 1
	24	Social Investigation and Discrimination	0hr and 1hr trial	Batch 1
	25	Social Investigation and Discrimination	24hr trial	Batch 1
	26			
	27			
	28 29			
	30	Social Investigation and Discrimination	Habituation	Batch 2
	31	Social Investigation and Discrimination	0hr and 1hr trial	Batch 2
April	1	Social Investigation and Discrimination	24hr trial	Batch 2
	2	Animal Cull		
	3	Animal Cull		
	4 5			
	6	Animal Cull		
	7	Animal Cull		
	8	Animal Cull		
	9	Animal Cull		

3.4.3 Open Field

The open field task (Hall 1951) measures the behaviour of a test animal in a novel environment. It is designed to measure locomotor activity, exploration and can also be used as a measure of anxiety. The task creates an environment where the animal's natural drive to explore conflicts with its avoidance of an unfamiliar arena. Typically mice engage in thigmotaxic behaviour (staying close to the walls of the arena) with reduced exploration of the open, central area of the open field.

The open field was a square white acrylic box with dimensions 72cm × 72cm × 33 cm. The animal was placed in the outer part of the arena facing an outer wall and allowed to freely explore the arena for 5 minutes. A video camera placed above the arena allowed movement to be tracked using an automated tracking system (Ethovision, Noldus Information Technologies). The number of faecal boli and urination were recorded at the end of the test. Concentric squares of equal distance from the periphery were defined in Ethovision as the 'outer', 'middle' and 'central' zones in order to determine the number of entries into, and time spent in, these zones of the arena. In addition, the latency to enter the inner zones as well as locomotor activity (distance moved and average speed) in all three zones of the arena was measured by the tracking system (*Figure 3.2*).

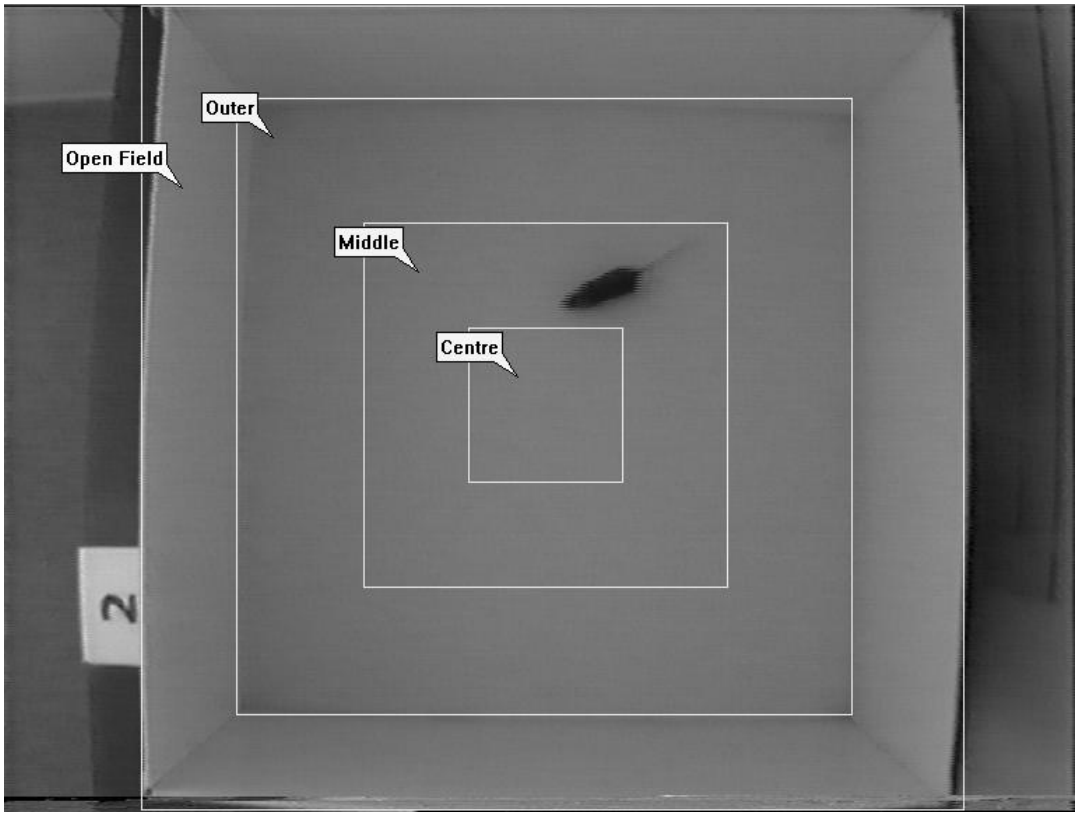


Figure 3.2 - Photograph of the Open Field Arena

The arena is divided in Ethovision into outer, middle and centre

3.4.4 Novel Object Exploration and Discrimination

The novel object task was used to determine short and long term memory by introducing animals to previously unseen objects and measuring the time they spend exploring the objects (Ennaceur and Delacour 1988). During the task the animals were exposed to three very different objects spread over three trials. The amount of time the animals spend sniffing and investigating the objects can be interpreted as a measure of exploration, and the amount of time spent with the different objects can be used as a measure of discrimination based on previous recall of an object. The same arena used in the open field task was used for the novel object task as mice had been exposed to this arena in the previous task, with the aim of maximising exploration of the objects and not the (now familiar) arena. Three pairs of different objects were used for this task. The three objects were yellow pots; china cups and metal tube were made of different materials (plastic, china and metal respectively) and differed in height, shape and colour. A video camera placed above the arena allowed

movement to be tracked using an automated tracking system (Ethovision, Noldus Information Technologies). The number of faecal boli and urination were recorded at the end of the test. The arena was divided virtually in Ethovision as in the open field task, see *Figure 3.3*.

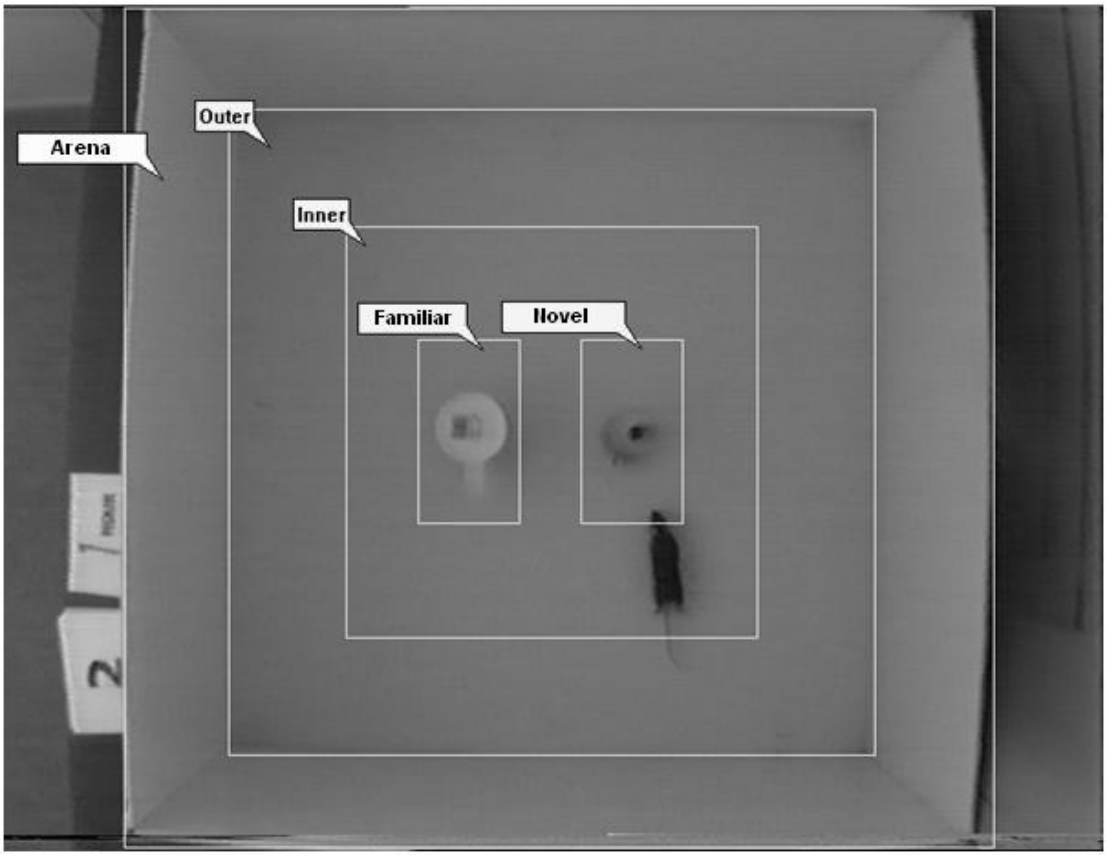


Figure 3.3 - Photograph of Novel Object arena

The arena is split in Ethovision into inner and outer and the areas around the objects into separate areas

The test animal was placed in the outer part of the arena facing an outer wall and allowed to freely explore the arena for 5 minutes. In the first trial, two identical, novel objects were placed 10 cm apart in the centre of the arena and the time spent exploring each object was hand coded using the Ethovision software. The second trial was carried out one hour later, designed to measure short-term memory and used one of the objects from the first trial (which was now a familiar object) and an unfamiliar object (the novel object). The third trial was carried out 24 hours after the second trial (to test long-term memory) using the novel

object from the second trial (now becoming the familiar object in this trial), and a new, unfamiliar object (*Figure 3.4*). Exploration of the object was defined as approaching the object and having physical contact; sniffing the object, turning the head towards it even for a brief period, touching it with the forepaws. Passive contact with the object (e.g. when the animal was sitting immobile but close to the object) was not considered to be object exploration. The order of object presentation and object location in the arena (left of right) were counterbalanced across the groups to minimize any potential confounds due to object preferences/biases.

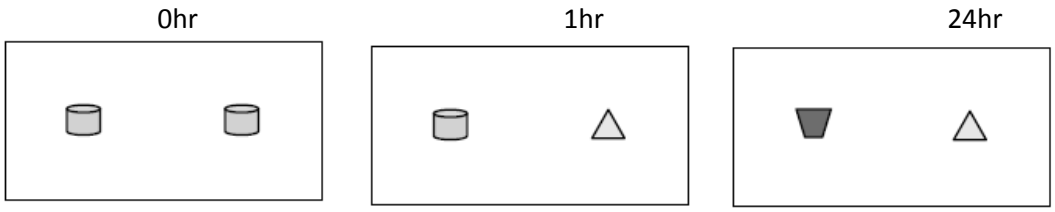


Figure 3.4 - Example of Object Configuration in the Three Trials of Novel Object

In the first trial, two identical objects are presented to the animal to investigate exploration. In the second trial (1hr later), an object from the first trial plus a novel object are placed in the arena. In the third trial (24hrs after the second trial) the novel object from the second trial and a new novel object are placed in the arena.

3.4.5 Holeboard

The holeboard test is used to measure activity and exploration in a novel arena (Boissier and Simon 1962; Boissier, Simon et al. 1964; Nolan and Parkes 1973). The number of times the mouse pokes its nose into the holes and the time spent nose poking is used as a measure of exploration. The Truscan Photo Beam Activity System (Coulbourn Instruments, Whitehall, PA) was used, which consists of an arena (25.4cm x 25.4cm) and a nose poke floor with 16 holes in a 4×4 array with sensor rings to track movement and nose poking. The beams were spaced 1.52 cm apart providing a 0.76 cm spatial resolution. Each mouse was placed in the arena and the movement, the number of, and the time spent, nose poking was recorded automatically by beam breaks for 5 minutes using the Truscan program.

3.4.6 Social Investigation and Discrimination

The social investigation and discrimination test is used to assess short and long term memory as in the novel object task, but also to investigate social interaction of a test mouse to a conspecific mouse. The social behaviour of the test mice towards a juvenile conspecific was assessed in a series of 5 minute trials (Winslow 2003). The test mouse was habituated in an arena (36cm ×20cm × 14cm) for 5 minutes, after which a male juvenile C57BL/6J conspecific (aged 4 weeks) was introduced for a further 5 minutes. Juvenile mice were used to minimise any aggression. During this trial, social behaviour including social sniffing (head and upper body of the conspecific), anogenital sniffing and allogrooming by the test mouse towards the conspecific was scored from videotape by an observer blind to the group factor of paternal age (*Figure 3.5*). Following/chasing and aggressive behaviours of the test mouse toward the conspecific mouse were also scored. One hour after the initial trial, the original familiar conspecific mouse and a second, novel conspecific mouse were introduced to the test mouse and behaviours were scored again (to test short-term memory). 24 hours after the second trial, a novel conspecific mouse was introduced to the test mouse as well as the novel mouse from the second trial, now the familiar animal (to test long-term memory). All behaviours were recorded and rescored for social behaviour at a later date and also rescored by an independent scorer to check for scoring consistency.

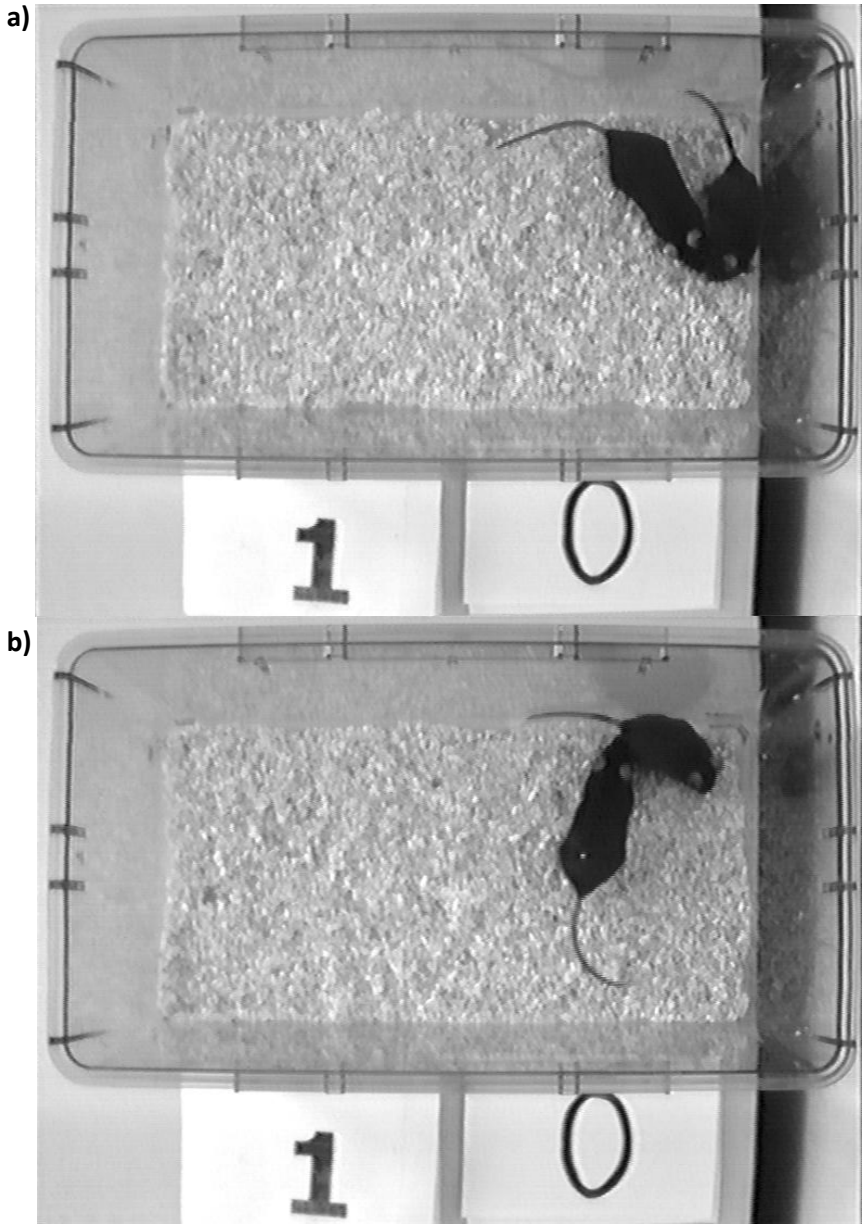


Figure 3.5 - Photograph of Social Behaviour Task

Smaller mouse with marking on tail is conspecific.

a) Sniffing

b) Anogenital sniffing

3.4.7 Statistical Analysis

Behaviours measured in the social interaction task, holeboard, novel object and open field were compared using a one-way ANOVA (main factor of paternal age) and Student's t-test, performed within R. Significance level was set at $p < 0.05$.

3.4.8 Dissection

Following behavioural testing, mice were culled by cervical dislocation and decapitated. Ovaries were removed from female breeders and testis and epididymis were removed from male breeders and offspring. Brains from the offspring were removed and the cerebellum, frontal cortex, striatum, hippocampus and hypothalamus were dissected out (*Figure 3.6*). The remaining brain tissue was also stored. Whole brains were removed from female and male breeders and half were flash frozen in dry ice and half stored in PFA/sucrose. All other tissues were flash frozen on dry ice and stored at -80^{o}C until extraction of nucleic acids (detailed in section 2.2)

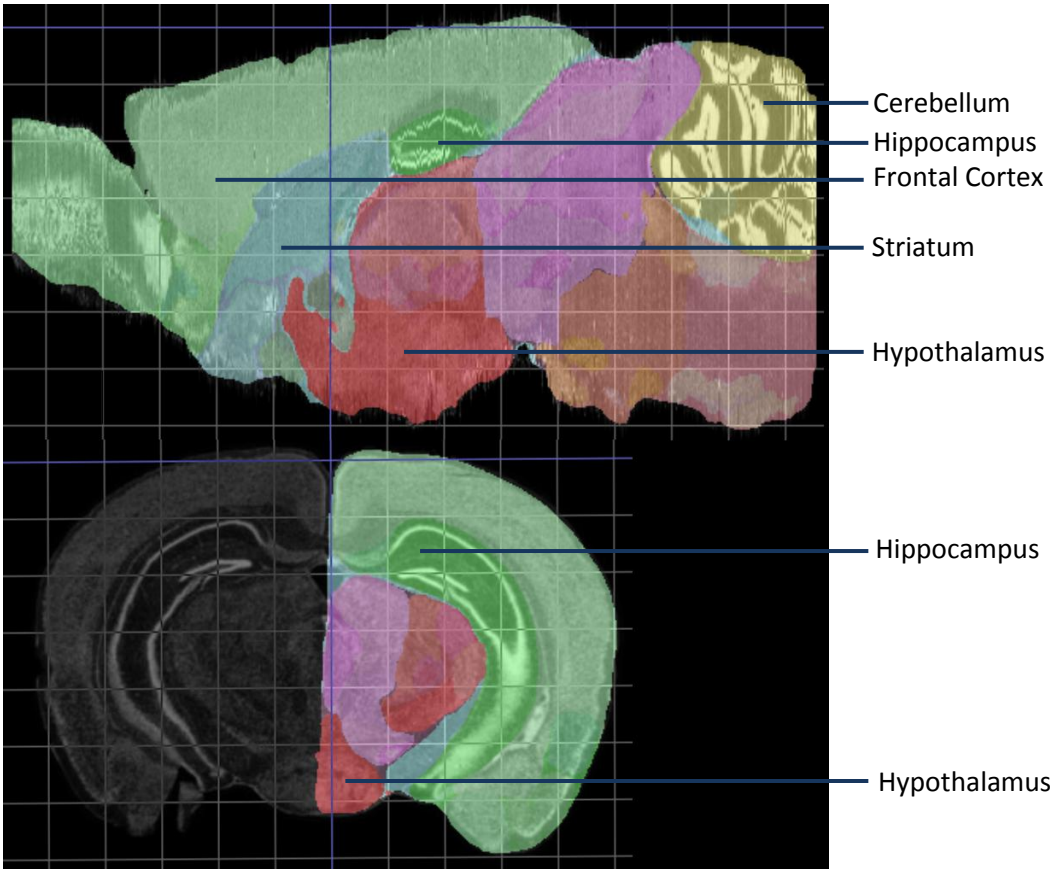


Figure 3.6 - Mouse Brain Maps

Diagram highlighting the regions dissected from mice in this study. Created using Allen Mouse Brain Atlas (Lein, Hawrylycz et al. 2007).

3.5 Results

For the graphs of the behavioural results presented in the section below, n = 12 for offspring of young males, n = 12 for offspring of old fathers and n = 9 for offspring of very old fathers. The results are presented separately for each of the three test groups (offspring of young, old or very old father) as well as for the offspring of young fathers (n = 12) versus a combination of the old and very old groups (fathers over 10 months old (n = 21)). Error bars on all graphs presented show standard error mean (SEM). * is used to signify a p-value below 0.05, + is used to signify a p-value between 0.1 and 0.05.

3.5.1 Body Weights

Data from both behaviourally-tested and test-naïve mice are presented below. No differences were observed in the weights of the offspring between groups at 8 weeks (F (2, 49) = 1.02, p = 0.37), 9 weeks (F (2, 49) = 1.10, p = 0.34) or 11 weeks (F (2, 49) = 1.44, p = 0.25) (*Figure 3.7*).

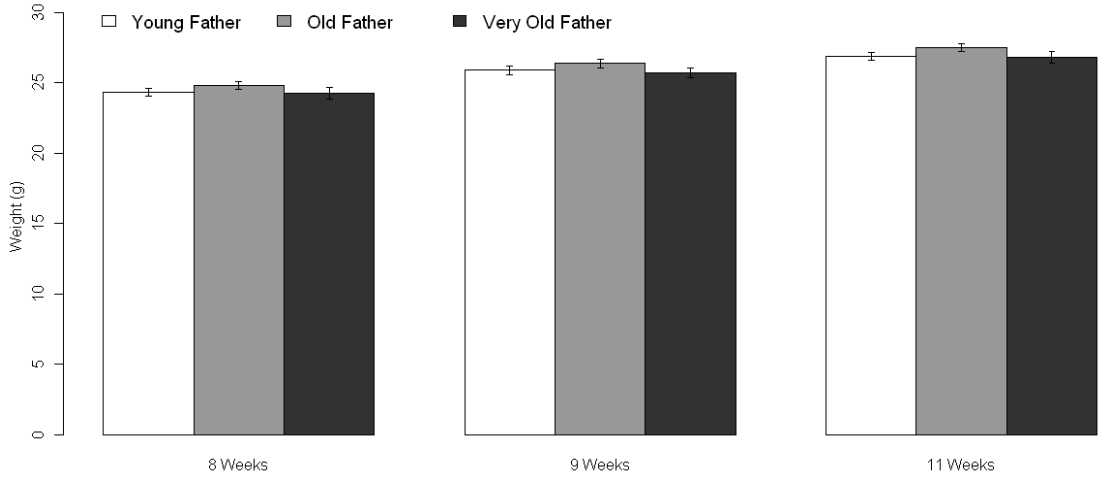


Figure 3.7 - Weights of Offspring by Group

Young Father n = 18, Old Father n = 23, Very Old Father n = 11

3.5.2 Open Field

Results of the open field task between the three groups are shown in *Figure 3.8*. Offspring of old fathers were significantly less exploratory in the open field task, taking longer to enter

the central zone of the arena than the offspring of young fathers (t = 1.7837, d.f. = 22, p = 0.04) (*Figure 3.8*c). There were no significant differences observed between the offspring of very old fathers and the offspring of young fathers (t = -0.81, d.f. = 19, p = 0.19). There were no significant differences in the time spent in the middle (F (2,30) = 0.57, p = 0.57) or central zones (F (2,30) = 0.88, p = 0.43) (*Figure 3.8*e), the frequency of entries into the middle (F (2,30) = 0.43, p = 0.65) or central zone (F (2,30) = 0.33, p = 0.72) (*Figure 3.8*d), the speed of movement (F (2,30) = 0.56, p = 0.58) (*Figure 3.8*b) or in the total distance moved during the task (F (2,30) = 0.39, p = 0.68) between any of the groups (*Figure 3.8*a).

Results of the open field task between the offspring of young fathers and offspring of fathers over 10 months are shown in *Figure 3.9*. When offspring of young fathers were compared to offspring of fathers over 10 months (combining the offspring of old and very old father groups), most of the results remain the same although the significance between the groups in the latency to enter the middle zone was no longer significant (t = -1.3493, d.f. = 31, p = 0.16) (*Figure 3.9*c).

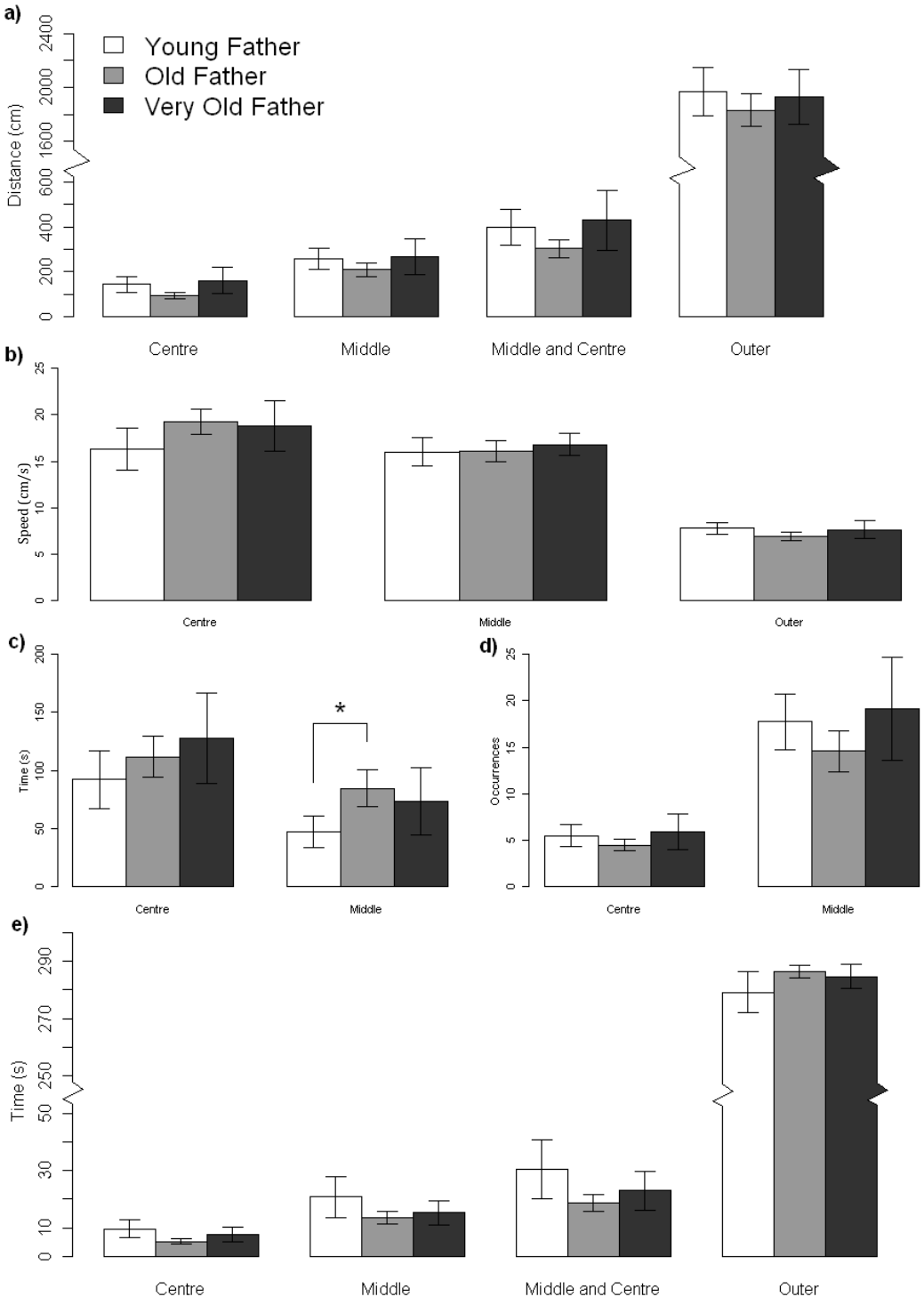


Figure 3.8 - Results of measures from the Open Field task

a) Total distance moved by group in each area of the arena

b) Average speed by group in each area of the arena

c) Latency to enter the two central areas of the arena by group

d) Number of times entering the two central areas of the arena by group

e) Time spent in each area of the arena by group

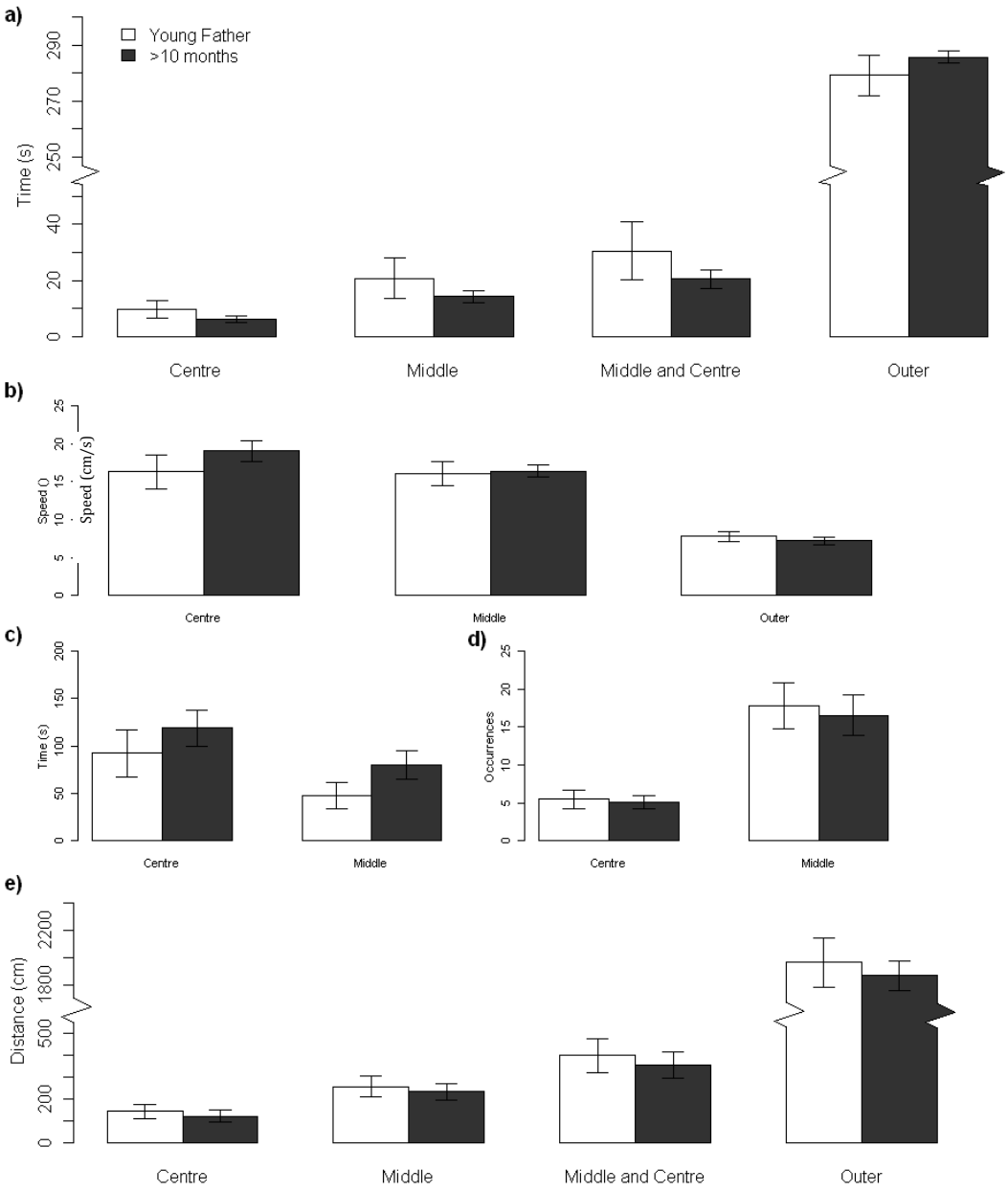


Figure 3.9 - Results of Open Field Task: Young vs. >10 Months Fathers

a) Total distance moved by group in each area of the arena

b) Average speed by group in each area of the arena

c) Latency to enter the two central areas of the arena by group

d) Number of times entering the two central areas of the arena by group

e) Time spent in each area of the arena by group

3.5.3 Novel Object Exploration and Discrimination

Results of the novel object task between the three groups are shown in *Figure 3.10*. No effect of paternal age was observed in the measures of exploration or locomotor activity in the novel object task (*Figure 3.10*a-f). In measures of discrimination, the offspring of old fathers were better at distinguishing between the novel and familiar object in the 1hr task as they spend more time exploring the novel object than the familiar one compared to the offspring of young fathers and very old fathers which did not discriminate between objects (*Figure 3.10*g&h). However, overall levels of exploration were very low in this task, especially during the 24hr trial when most of the animals did not approach the objects at all (data not shown). Therefore it is difficult to draw any conclusions from the data except for the locomotor activity measures (distance moved) where reasonable levels of activity were observed but there were no significant differences between groups (*Figure 3.10*a).

Results of the novel object task between the offspring of young fathers and offspring of fathers over 10 months is shown in *Figure 3.11*. When the offspring of young fathers were compared to offspring of fathers over 10 months the results are the same as those from the three group comparison. Again the fact that the animals did not spend much time exploring the objects means that conclusions cannot be drawn from this data except to say there were no differences in locomotor activity between the two groups (*Figure 3.11*a) as in the comparisons between the three groups.

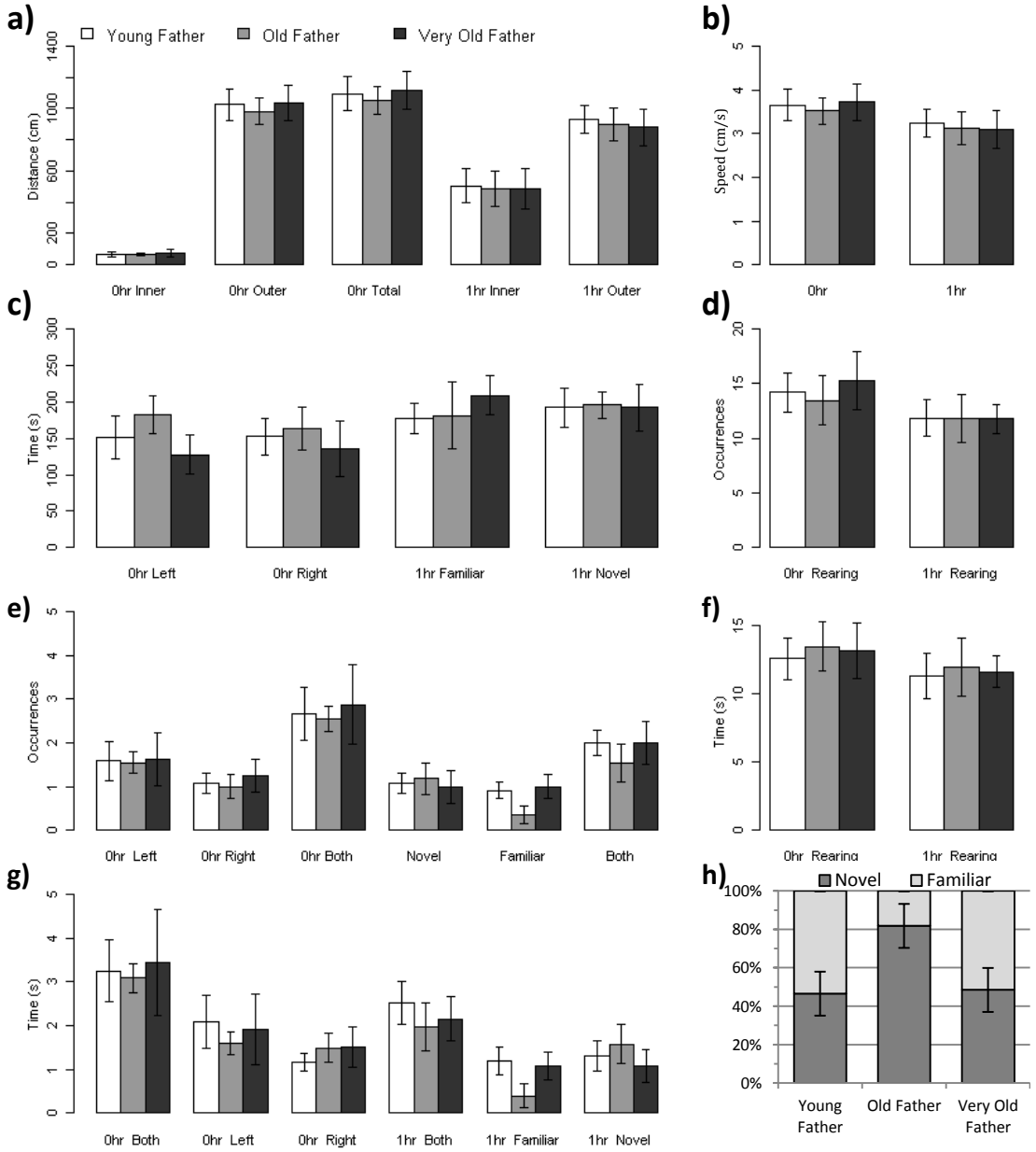


Figure 3.10 - Results of the Novel Object Task

a) Distance moved in areas of arena in 0hr and 1hr task

b) Average Speed during 0hr and 1hr tasks

c) Latency to approach objects in 0hr and 1hr tasks

d) Frequency of rearing in 0hr and 1hr tasks

e) Frequency of object visits during 0hr and 1hr tasks

f) Time spent rearing during 0hr and 1hr tasks

g) Time spent with objects during 0hr and 1hr tasks

h) Percentage of time spent with objects in 1hr task

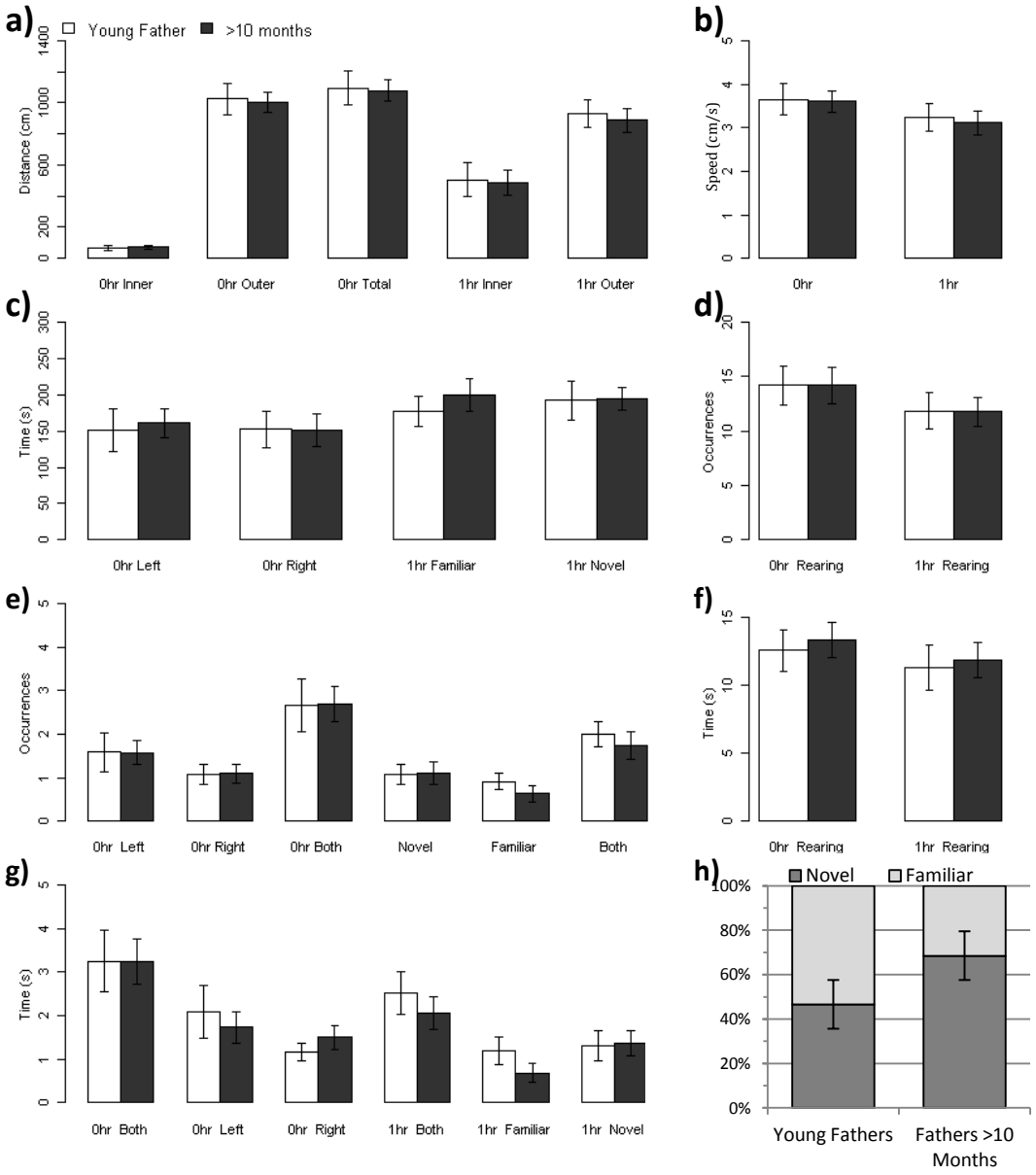


Figure 3.11 - Novel Object Task Results: Young vs. Fathers >10 Months

a) Distance moved in areas of arena in 0hr and 1hr task

b) Average Speed during 0hr and 1hr tasks

c) Latency to approach objects in 0hr and 1hr tasks

d) Frequency of rearing in 0hr and 1hr tasks

e) Frequency of object visits during 0hr and 1hr tasks

f) Time spent rearing during 0hr and 1hr tasks

g) Time spent with objects during 0hr and 1hr tasks

h) Percentage of time spent with objects in 1hr task

3.5.4 Holeboard

The results of the holeboard task in the three groups are shown in *Figure 3.12*. Offspring of old fathers demonstrated reduced exploration in the holeboard task compared to the offspring of young fathers, making fewer nose pokes (t = 2.29, d.f. = 22, p=0.03) (*Figure 3.12*b) and spending less time nose poking (t = 2.3, d.f. = 22, p = 0.03) (*Figure 3.12*c). No significant differences were seen between the offspring of very old fathers and the other two groups. No significant differences were evident in the total distance moved (F (2, 30) = 1.25, p = 0.30) (*Figure 3.12*a) or centre entries (F (2, 30) = 0.14, p = 0.86) in the holeboard arena between the three groups (*Figure 3.12*d).

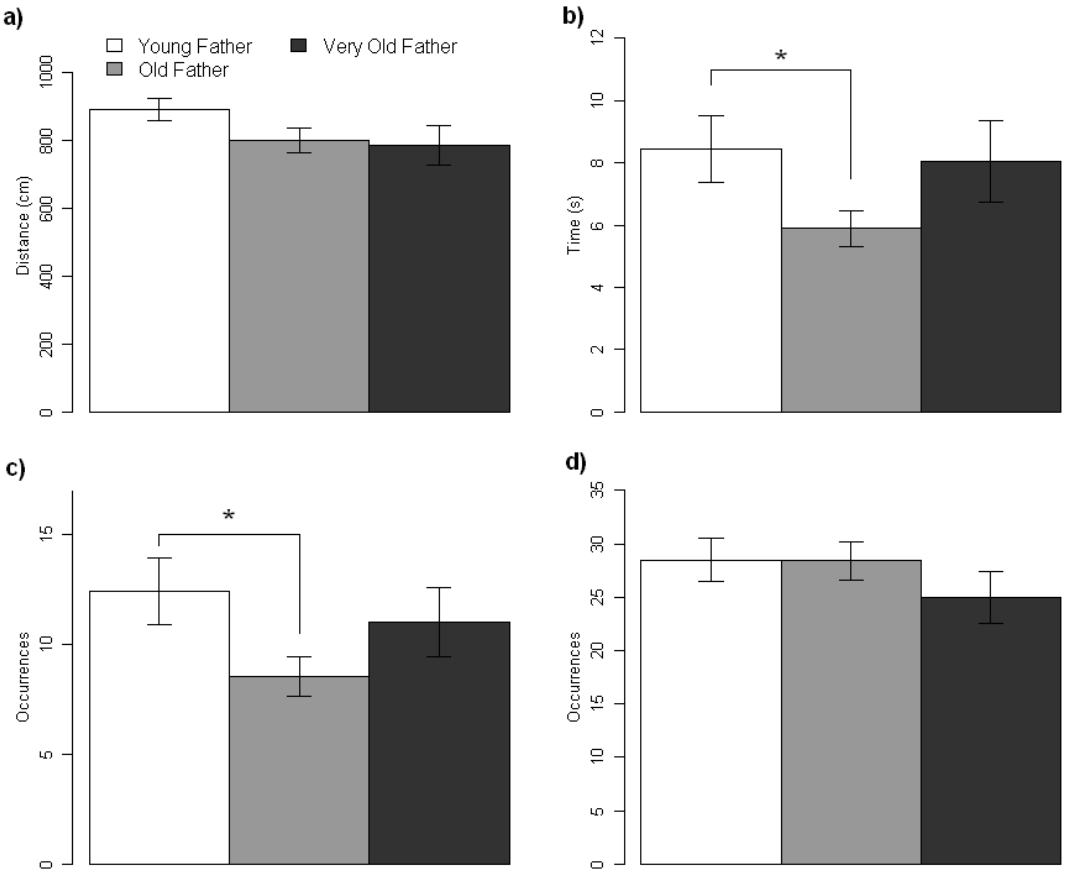


Figure 3.12 - Results of Holeboard Task

a) Total distance moved during task by group

b) Time spent nose poking by group

c) Total number of nose pokes by group

d) Total centre entries by group

The results of the holeboard task between the offspring of young fathers and offspring of fathers over 10 months is shown in *Figure 3.13*. The offspring of fathers over 10 months performed significantly fewer nose pokes (t = 1.96, d.f. = 31, p = 0.04) (*Figure 3.13*c). There was also a trend for significance in the total time spent nose poking (t = 1.65, d.f. = 31, p = 0.07) between the groups (*Figure 3.13*b). The distance moved (t = 1.38, d.f. = 31, p = 0.14) (*Figure 3.13*a) and entries into the centre (t = 0.39, d.f. = 31, p = 0.66) remain non-significant (*Figure 3.13*b).

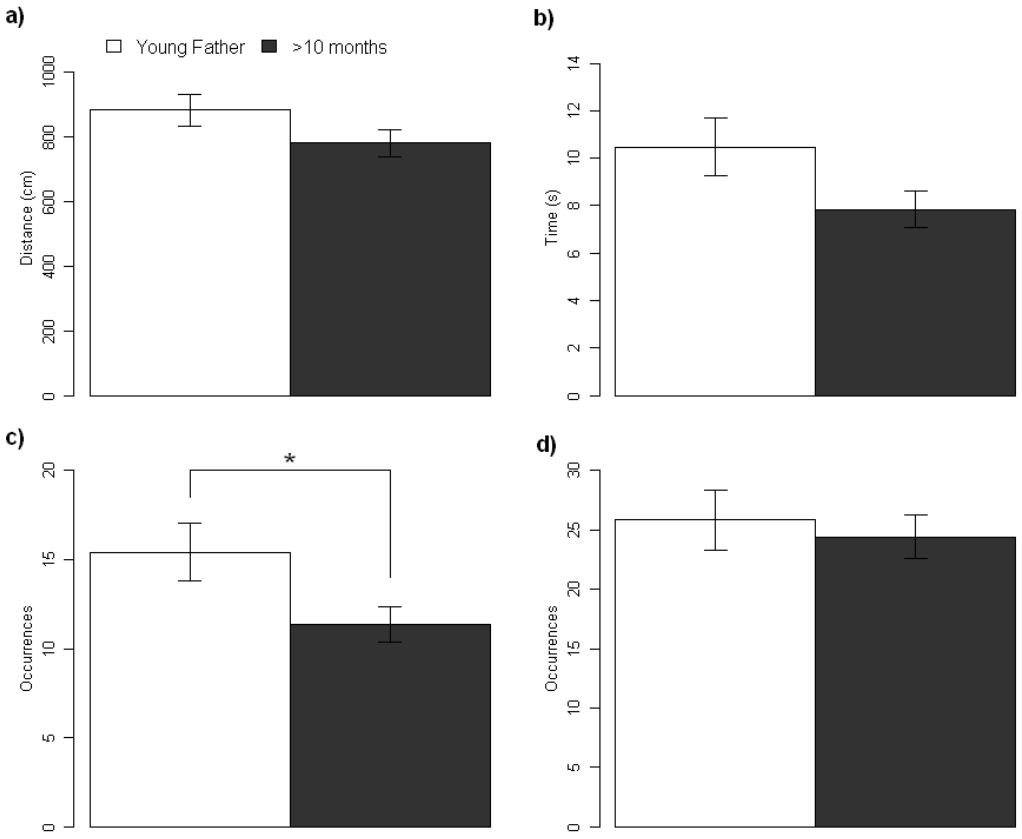


Figure 3.13 - Results of Holeboard Task: Young vs. Fathers >10 Months

a) Total distance moved during task by group

b) Time spent nose poking by group

c) Total number of nose pokes by group

d) Total centre entries by group

3.5.5 Social Investigation and Discrimination

The results of the social investigation task for the three groups are shown in *Figure 3.14*. The offspring of old fathers engaged in significantly less overall social investigation than the offspring of young fathers, ($t = 2.23$, $d.f. = 22$, $p = 0.03$) (*Figure 3.14*d). The main differences were seen in social sniffing between offspring of young fathers and offspring of old fathers ($t = 2.38$, $d.f. = 22$, $p = 0.03$) (*Figure 3.14*a). This is the main contributor to the overall reduction in total time spent in social behaviour. There were no significant differences in overall locomotor activity. No significant differences were seen in the time spent engaging in total social behaviour between the offspring of very old fathers and the offspring of young fathers ($t = 1.69$, $d.f. = 19$, $p = 0.14$), and offspring of old fathers ($t = -0.64$, $d.f. = 19$, $p = 0.52$). Even though the differences between the offspring of young fathers and offspring of very old fathers are not significantly different, the total time in social behaviours is still reduced in the offspring of very old fathers (*Figure 3.14*d). Values from the second scorer for the total time spent in social behaviour was also significant ($t = 2.19$, $d.f. = 22$, $p = 0.03$), although the second scorer did not split the behaviours by different type.

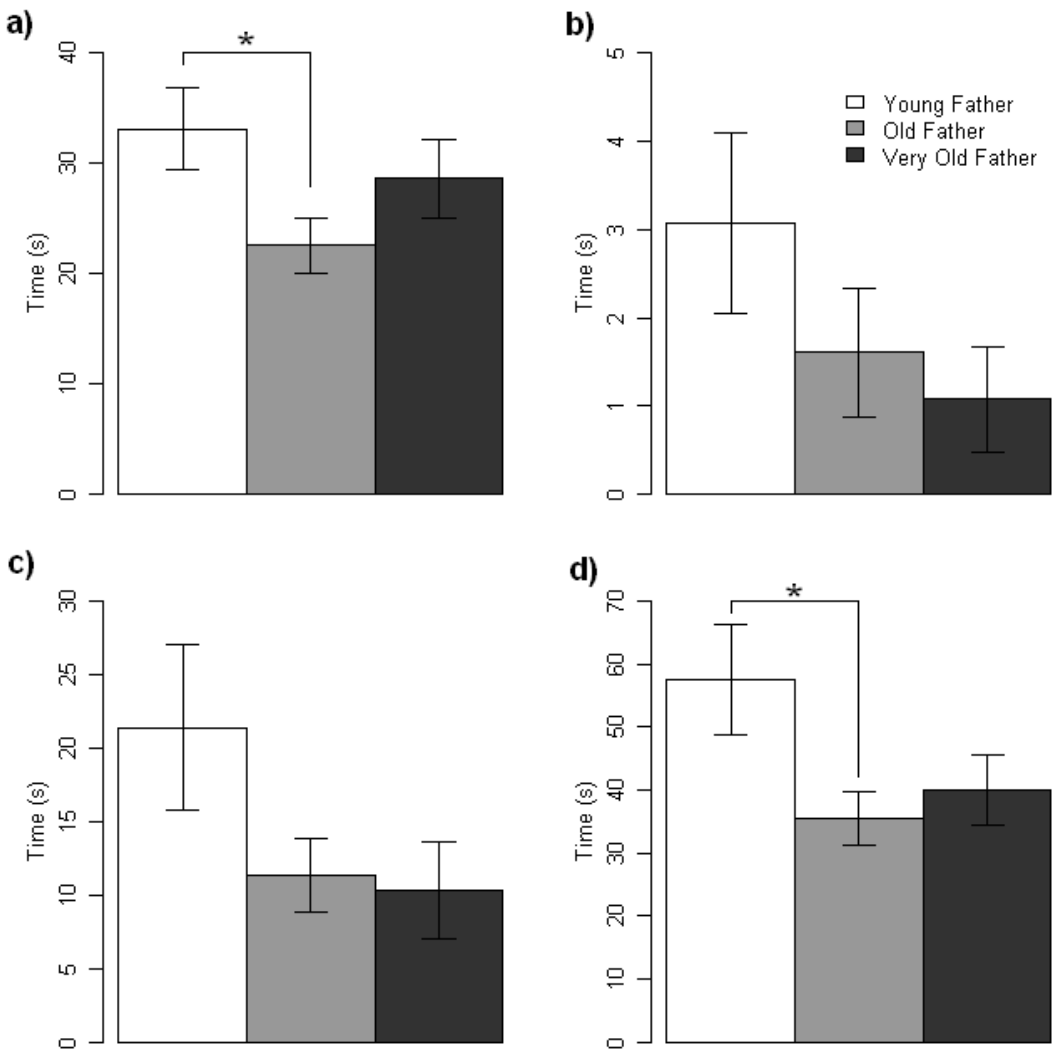


Figure 3.14 - Results of the Social Investigation Task

a) Time spent in social sniffing behaviour

b) Time spent allogrooming

c) Time spent in anogenital sniffing behaviour

d) Total time spent in all social behaviours

Social discrimination in the 1hr and 24 hr task between the three groups is shown in *Figure 3.15*. In the 1hr trial, offspring of old fathers spent a significantly larger percentage of time than the offspring of young fathers with the novel conspecific mouse than the familiar mouse (*Figure 3.15*a) ($t = -2.51$, d.f. = 22, $p = 0.02$), but there was no significant discrimination during the 24hr trial (*Figure 3.15*b) ($t = 0.445$, d.f. = 22, $p = 0.66$). There were no differences between the offspring of very old fathers and the offspring of young fathers in either the 1hr ($t = -1.41$, d.f. 19, $p = 0.15$) or 24hr trial ($t = 0.22$, d.f. = 19, $p = 0.82$).

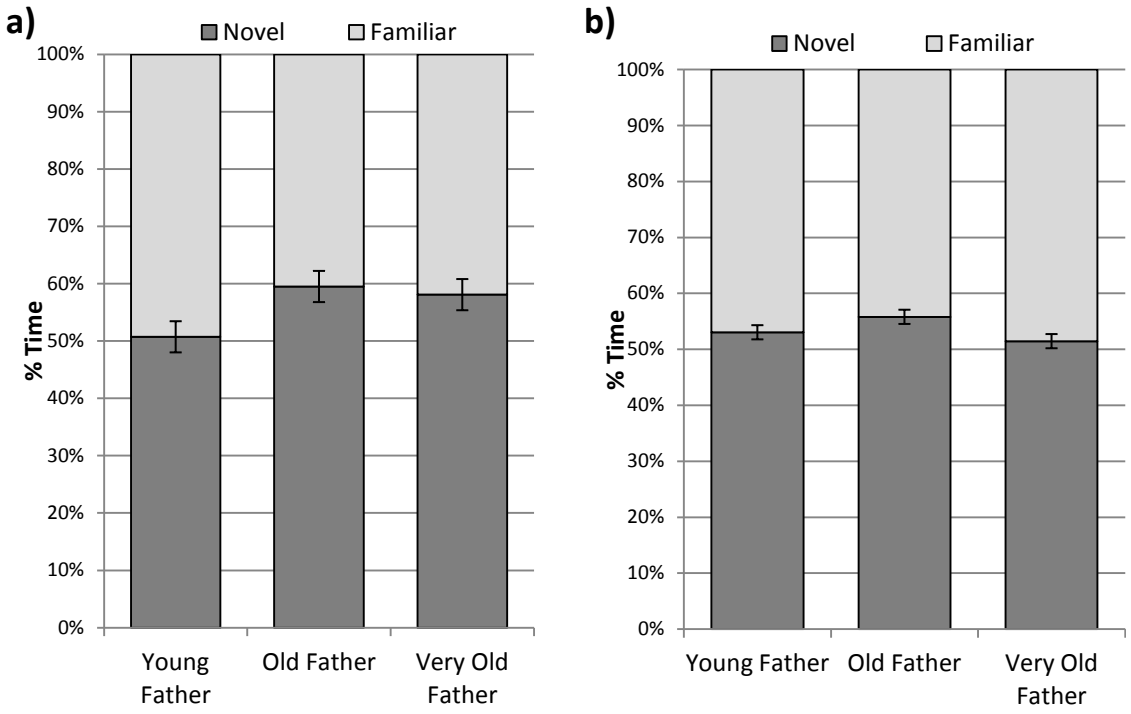


Figure 3.15 - Conspecific Discrimination in 1hr and 24hr Social Behaviour Task

a) 1hr after original social investigation task

b) 24hrs after original social investigation task

The results of the social behaviour task between the offspring of young fathers and offspring of fathers over 10 months is shown in *Figure 3.16*. When comparing the offspring of young fathers with offspring of fathers over 10 months, a similar result to the previously shown analysis is observed. The offspring of fathers over 10 months spend significantly less time in social sniffing behaviour (*Figure 3.16*a) ($t = -0.29$, d.f. = 31, $p = 0.05$) and anogentical sniffing (*Figure 3.16*c) ($t = -1.68$, d.f. = 31, $p = 0.04$) and less overall time in social activity (*Figure 3.16*d) ($t = 2.05$, d.f. = 31, $p = 0.02$) than the offspring of young fathers. A trend towards a significant difference was seen in time spent in time spent in allogrooming behaviour between the offspring of young fathers and offspring of fathers over 10 months ($t = 1.41$, d.f. = 31, $p = 0.10$) (*Figure 3.16*b).

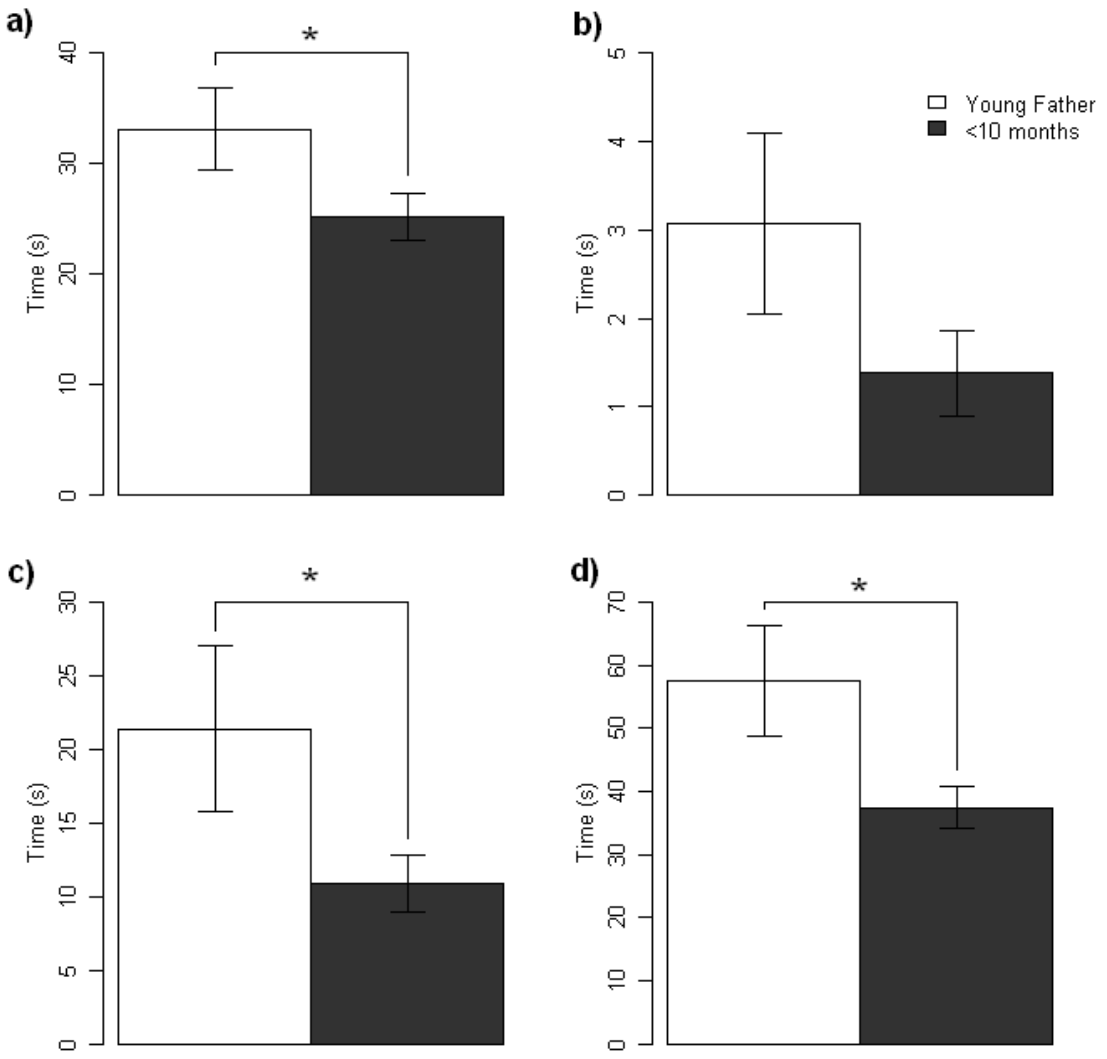


Figure 3.16 - Results of Social Investigation Task: Young vs. Fathers >10 Months

a) Time spent in social sniffing behaviour

b) Time spent allogrooming

c) Time spent in anogenital sniffing behaviour

d) Total time spent in all social behaviours

Social discrimination in the 1hr and 24 hr task between the offspring of young fathers and offspring of fathers over 10 months is shown in *Figure 3.17*. The offspring of fathers over 10 months discriminated between the novel and familiar mouse in the 1hr trial better by spending more time with the novel mouse than the famiiar mouse compared to the offspring of young fathers (*Figure 3.17*a) ($t = -2.13$, d.f. = 31, $p = 0.03$), but again there was no significant discrimination in the 24hr trial (*Figure 3.17*b) ($t = 0.14$, d.f. = 31, $p = 0.88$).

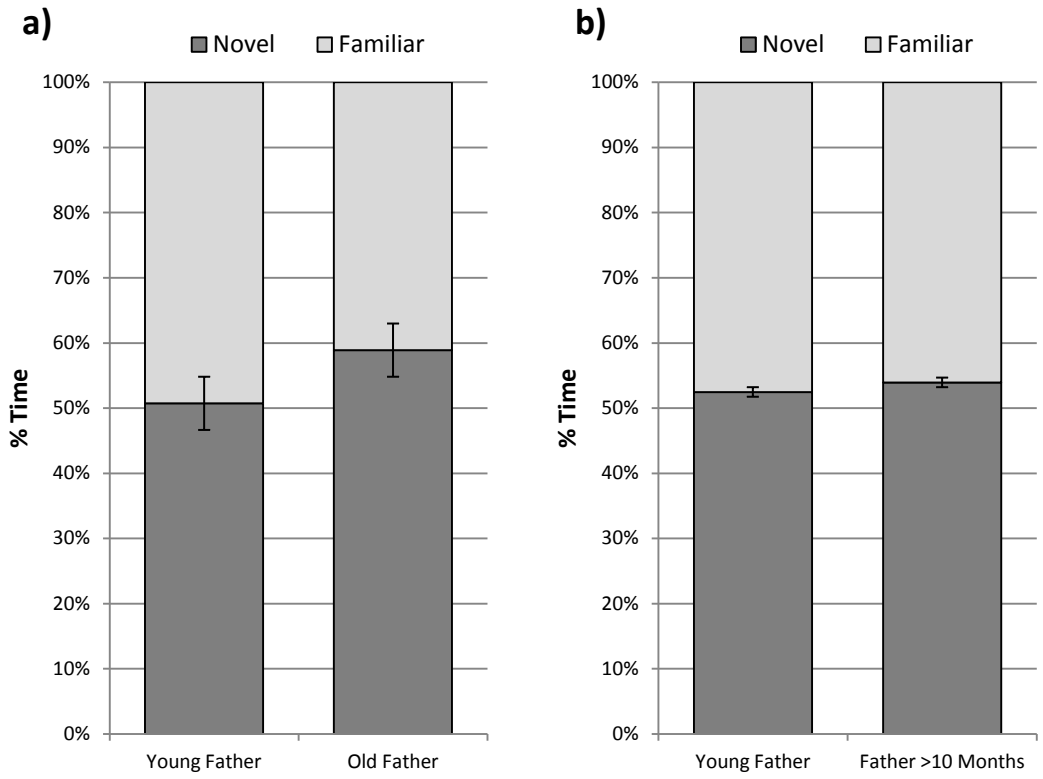


Figure 3.17 - Conspecific Discrimination in 1hr and 24hr Social Behaviour Task: Young vs. Fathers >10 months

a) 1hr after original social investigation task

b) 24hrs after original social investigation task

3.6 Discussion

3.6.1 Behavioural Outcomes of Advancing Paternal Age

In this chapter, an investigation into whether advancing paternal age is associated with changes in behaviour in the offspring. The primary observation is that the offspring of older fathers engaged in significantly less social and exploratory behaviour than the offspring of young fathers. Significantly less social behaviour was seen in the offspring of fathers over 10 months when combining the offspring of old and very old fathers, compared to the offspring of young fathers. Locomotor measures in all tests were similar between all three groups in the open field, novel object, holeboard and social behaviour tasks. Behavioural changes were limited to time spent in social interaction in the social behaviour task, latency to explore in the open field and time spent nose poking and number of nose pokes in the holeboard task. Given

that other factors were controlled, paternal age is likely to be the cause of these behavioural outcomes most significantly in social behaviours. This observation is particularly interesting given the known social deficits observed in ASD, schizophrenia and BD as discussed at the beginning of this chapter and in section 1.3.

Differences between offspring of very old fathers and the offspring of young fathers were not observed as consistently as in the offspring of old fathers, although in many cases the direction of effect was the same. In the social behaviour task the offspring of very old fathers did not engage in significantly less or more social behaviour than the offspring of young fathers ($t = 1.68$, $d.f. = 19$, $p = 0.13$), or old fathers ($t = 0.64$, $d.f. = 19$, $p = 0.52$) although when combined with the offspring of old fathers, the reduction was significant. One reason for the difference in the offspring of old fathers and not very old fathers could be due to the smaller number of animals in this group. The very old father group has 25% less animals than the young father and old father groups and so the power to detect differences in this group is reduced. The smaller group size is due to less litters being born to the very old fathers. However, these observations do partly fit with the previous human literature. For example in ASD, where increasing paternal age increases the risk of developing autism until age 44 when the risk levels out and even drops slightly at age 50-59 and then increases again after age 60 (Grether, Anderson et al. 2009) (*Figure 3.18*). In order to test this further, more groups of breeders of older than 12 months would need to be used.

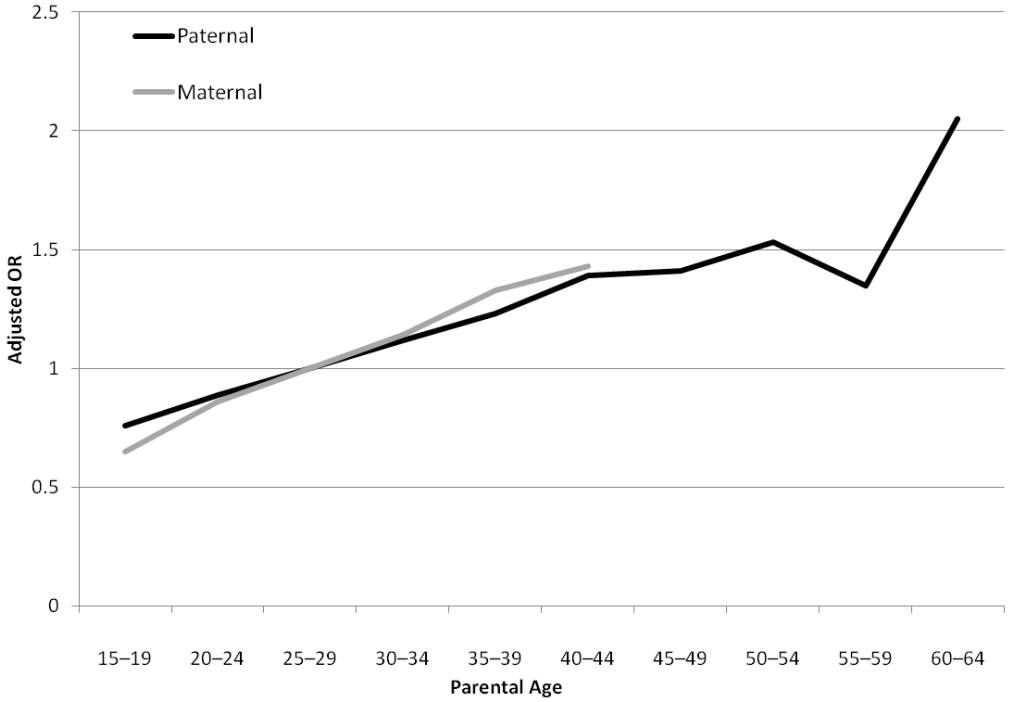


Figure 3.18 - Parental Age and Autism Risk

Maternal and Paternal age and risk of autism in California residents births 1989-2002

Adapted from Grether et al (2009) (Grether, Anderson et al. 2009)

A second reason for the less severe behavioural changes between the offspring of very old fathers and the offspring of young fathers could be due to differences in the father's environment. The very old fathers were purchased from Charles River at a separate time and could possibly be part of a different batch of animals from the other two groups. This could mean there are very slight differences in genetics due to CNVs or point mutations established at the beginning of the breeding line. Whether the behaviour was caused by environment, group size, and small genetic variations or as a parallel to the evidence from some studies is not clear.

It is unfortunate that the novel object task appeared to be not successful in the current study. The animals in all groups displayed limited exploration of the objects and during the 24hr trial very few animals approached the objects at all. The C57BL/6J strain has previously been shown to perform well in this task but this could be a consequence of single housing. It has previously been shown that single-housed mice perform worse in the novel object task

than groups housed animals (Voikar, Polus et al. 2005) although this only affected the discrimination ratio and not the total amount of time exploring the objects. As the novel object task was unsuccessful, there was only one measure of short and long term memory from the social behaviour task but this task was more designed to look at social behaviour rather than memory and so having a purely memory based task would have been more informative. Also as we saw a difference in social investigation, there is a potential confound in the use of the social investigation for a memory task.

3.6.2 Animal Studies of Paternal Age

Although there are currently few comparable studies looking at the effects of paternal age on the behaviour of rodents, some comparisons can still be made with the results of this study. A summary of results across all rodent studies of paternal age to date is given in (*Table 3.3*). The first rodent study to look at the behavioural effects of paternal age showed that offspring of older fathers in rat (6 months) had decreased learning capacity than the offspring of standard age rats (2.5 months), and that the offspring performed worse in these task as paternal age increased further (10-23 months) (Auroux 1983).

In a separate study looking at developmental and behavioural effects of advanced paternal age in mice (Garcia-Palomares, Pertusa et al. 2009), offspring of fathers of 120 weeks showed a decreased ability to self right on postnatal day 4-6 but this effect was lost by postnatal days 7-10. The offspring in this group also had reduced spontaneous motor activity in a 60 minute trial and reduced passive-avoidance learning. The authors observed no differences between the offspring of their standard breeding age group of 12 weeks and the offspring of fathers of 70 weeks and 100 weeks. It is difficult to draw direct comparisons with the present study due to the very different ages of the breeding males and the tasks used, but the consistent differences between offspring of standard and old breeders provides further evidence for a phenotypic difference caused by advanced paternal age.

Foldi *et al's (Foldi, Eyles et al. 2010)* study looking at advanced paternal age in mice is the most similar to this study but differs in some methodology. Their results showed that the

offspring of old fathers spend less time on the open arms in the elevated plus maze indicating they were more anxious than the offspring of younger fathers. The offspring of old fathers were also more exploratory in the holeboard task and spent less time immobile in the forced swim task (Foldi, Eyles et al. 2010) where animals are put in a narrow cylinder of water with no escape and the amount of time spent immobile is recorded. This task has been proposed to measure behaviours related to depression. As the offspring of the advanced paternal age group were less immobile, it was suggested they experienced reduced depression-like behaviour. However, when split by sex, the effect was only seen in females and was not significant in males.

They also observed no differences between the offspring of young fathers and advanced age father group in the social behaviour task. This could be due to a number of reasons. First, the method of measuring social behaviour used in Foldi *et al's* study did not permit direct physical contact as the conspecific animals were separated from the test animal by perforated, transparent dividers. Second, the two ages they selected for their breeders were 4 months for the standard groups and 12-18 months for the old group. As previously discussed, weaker effects were observed in our study in the very old father group which was the closest in age to Foldi's advanced paternal age group. In the Foldi *et al* study, the main behavioural effects were seen in female mice and therefore direct comparisons with our study cannot be made as only male mice were tested. Even though the effect of advanced paternal age was only seen in female offspring, it still provides extra evidence for the hypothesis that advanced paternal age has an effect on behaviour.

Table 3.3 - Summary of Paternal Age Studies

	(Smith, Kember et al. 2009) (this study)	**(Auroux 1983)**	**(Garcia-Palomares, Pertusa et al. 2009)**	**(Foldi, Eyles et al. 2010)**
Animal	• Mouse (C57BL/6J)	• Rat (Wistar)	• Mouse (C57BL/6JIco)	• Mouse (C57BL/6J)
Sex Tested	• Males	• Males and Females	• Males and Females	• Males and Females
Father's Ages	• 2 months • 10 months • 12 months	• 2.5 months • 6 months • 10 months • 14 months • 18 months • 22 months	• 12 weeks • 70 weeks • 100 weeks • 120 weeks	• 4 months • 12-18 months
Tests	• Open Field • Novel Object • Holeboard • Social Behaviour	• Spontaneous Activity • Learned Avoidance	• Righting Reflex • Spontaneous Motor Activity • Passive-Avoidance	• Elevated Plus Maze (EPM) • Holeboard • Social Behaviour • Forced-Swim • Prepulse Inhibition

Table 3.3 Continued

	(Smith, Kember et al. 2009)	(Auroux 1983)	(Garcia-Palomares, Pertusa et al. 2009)	(Foldi, Eyles et al. 2010)
Results	• Increased latency to middle in Open Field (10 months) • Reduced exploration in Holeboard (10 months) • Reduced overall social interaction in Social Behaviour task (10 and 12 months)	• Reduced learning capacity with increasing paternal age	• Decreased ability to self right on postnatal day 4-6 in 120 week group • Reduced spontaneous motor activity for 120 week group • Reduced passive-avoidance learning in 100 and 120 week group	• Increased time in open arms in EPM (females only) • Increased exploration in Holeboard (females only) • Less time immobile in Forced-swim task (females only)

3.6.3 Limitations

There are several limitations to our study. First, only three different ages of male breeders were used making it difficult to determine whether the paternal age related affects on behaviour on linear or threshold dependant. As observed in some other studies discussed, having groups of increasing paternal age would allow the observation of potentially more extreme changes in behaviour and give an indication of breeding ages that are critical for the development of more severe phenotypes. Also the offspring of very old male group was significantly smaller than the other two groups and sired by a different batch of fathers. Therefore, the strongest conclusions should be drawn from the comparison between the offspring of young and old fathers. As different females were used for creating the offspring for the different groups we cannot sully rule out the effect of maternal care on the cause of behavioural changes. Variations in maternal care have been previously shown to affect behaviour in the offspring (Curley, Champagne et al. 2008) although this is unlikely to be an issue in the current study as litter effects were not observed.

The strongest result was in social behaviour but only one measure of social behaviour was used in the current study. As shown in Foldi *et al's* study, there are a number of ways to carry out a social behaviour or social preference test. Utilising the Y maze and dividers is one method (Foldi, Eyles et al. 2010) of measuring which animal the test mouse spends more time investigating. Also separating the arena into sections using dividers and holding conspecific mice under a round wired cage and measuring the time the test mouse spends in each area of the arena (Moy, Nadler et al. 2004). Allowing direct physical contact between mice in the social discrimination task could be a limitation if any aggression occurs and perhaps even the potential for an aggressive encounter may result in more avoidance behaviour in the test mice. Although significant differences were observed in the open field, holeboard and social behaviour tasks, a large number of tests were performed and so there is an issue of multiple testing. Since the significance level was set at 0.05, we would expect 5% of all findings to occur by chance. 156 tests between the three groups in all tasks were observed and so 7-8 of all values could be by chance.

As the only learning tasks included the behavioural battery (novel object) did not work, it is not possible to speculate as to whether there may be any learning deficits or memory

related to advancing paternal age in this study as shown in other studies looking at paternal age (Auroux 1983; Garcia-Palomares, Pertusa et al. 2009). Running the 24hr discrimination using the objects/conspecifics from the 1hr session could be a confound as requires the mice to forget the first set of objects/conspecifics and ideally this test should include an extra 0 hr session between the 1hr and 25 hr trials.

Only male offspring were used in the study due to the sex bias in autism where the ratio of male to female sufferers is 4:1 although as previously discussed this ratio drops with advancing paternal age (Anello, Reichenberg et al. 2009). However, we may be missing more convincing phenotypic changes associated with paternal age by not using females as there is evidence for female related social deficits in other studies (Foldi, Eyles et al. 2010). Paternal age effects are also seen in schizophrenia and BD as previously discussed and in these disorders the sex ratios are more balanced. Schizophrenia has a male to female ratio of 1.42:1 (Aleman, Kahn et al. 2003) and in BD the ratio is roughly equal (Robins, Helzer et al. 1984).

3.6.4 Conclusion

In conclusion, the behavioural data presented here provides evidence for the role of paternal age in the development of social deficits in resulting offspring. Taking into account other rodent paternal age studies and epidemiological finding, it is possible to conclude that paternal age is responsible for phenotypic changes in the offspring. The mechanism behind this association is unclear, but may provide important insight into the causes of several psychiatric disorders. Understanding the molecular changes associated with advanced paternal age is the focus for the remainder of my thesis.

Chapter 4 - Genome Wide Analysis of Copy Number Variants in a Mouse Model of Advanced Paternal Age

4.1 Abstract

De novo copy number variations (CNVs) are speculated to be a potential mechanism mediating the observed epidemiological association between advanced paternal age and the risk of developing neuropsychiatric disorders including ASD, schizophrenia and BD. Older males have been previously shown to have a larger incidence of CNVs and other chromosomal abnormalities in their sperm than younger males, in both humans and mice. As CNVs are aetiologically-associated with ASD, schizophrenia and BD, the increased rate of CNVs in older fathers may also mediate the association between advanced paternal age and disease risk. In this study, the offspring of young (2 month old) and old (>10 months old) fathers were analysed for CNVs using CGH on high-resolution genome-wide Nimblegen microarrays. The male and female breeders were also assessed in an attempt to differentiate inherited CNVs from *de novo* CNVs. It was observed that the incidence of inherited or *de novo* CNVs, their size, and the number of genes affected does not appear to be associated with advanced paternal age. Of note, however, the overall burden of CNVs was larger than anticipated indicating either that the rate of CNVs is larger than initially thought in inbred mouse strains or that methodological differences result in different estimates across studies.

4.2 Introduction

4.2.1 Copy Number Variation Background

As discussed previously in section 1.5, CNVs are defined as polymorphic deletions or duplications between 1Kb and 5Mb in size (Freeman, Perry et al. 2006), representing one of the major types of structural variation in the genome. An analysis of 270 individuals from four different ancestral populations found that 12% of their combined genome was covered by the observed CNVs (Redon, Ishikawa et al. 2006), and approximately 0.4% of inter-individual genomic variation in the population is estimated to be due to CNVs (Kidd, Cooper et al. 2008). Two unrelated individuals from the same population are estimated to differ in CNVs by between 5.9 and 6.3Mb, accounting for approximately 0.2% of the genome (McCarroll, Kuruvilla et al. 2008). CNVs can be either inherited (when one parent has an existing CNV which is transmitted on to their offspring), or *de novo* (where the CNV occurs spontaneously during gametogenesis or the development of the offspring and is not present in either

parent). *De novo* CNVs can occur by a variety of mechanisms, which are discussed later in this section.

CNVs can be represented by either duplications (where the genome contains one or more extra copies of a specific region of the genome) or deletions (where a region of the genome is represented by fewer copies than expected, i.e. less than two on autosomal chromosomes). A related type of polymorphism are 'copy neutral variations' such as 'inversions' (where a region of the genome is present as expected but is inverted), and 'translocations' (where a region of the genome is present as expected but has been moved to another chromosomal locus). More complex rearrangements, including combinations of the above mentioned types, are also possible. Some genes have been found to be highly variable in copy number, being present in the genome at between zero and 13 copies (Dalen, Dahl et al. 1998).

In humans, it has been hypothesised that there could be as many as 100 new point mutations per generation (Drake, Charlesworth et al. 1998). Although most of these are likely to occur in areas of non-coding DNA or be synonymous, some are likely to occur in coding regions and be non-synonymous but non-lethal so that they are introduced into the population. The frequency of *de novo* CNV events was traditionally suspected to be less frequent than the occurrence of point mutations, however this has been shown to be potentially untrue based on the incidence of disease-related rearrangements. The rate of point mutations is estimated to be 0.5–3.7 x 10^{-8} per nucleotide per generation with the combined rate of all mutations to be on average 1.8 x 10^{-8} per nucleotide per generation (Kondrashov 2003). Based on the prevalence of genomic disorders caused by known rearrangements, such as DiGeorge syndrome, Williams-Beuren syndrome, Smith-Magenis syndrome, Prader-Willi syndrome and Angelman syndrome, the estimate of *de novo* CNV rates is 1×10^{-4} in the population, and is thus potentially much higher than that of *de novo* point mutations in individual human autosomal dominant traits (Inoue and Lupski 2002).

In a recent study characterising the occurrence of *de novo* CNVs in humans, nine CNVs were observed in 722 transmissions, equating to a rate of 1/80 CNVs per individual (Itsara, Wu et al. 2010). The rates are also likely to be higher than previously speculated due to the

number of CNVs recently described in both disease free cohorts of humans (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004; Sharp, Locke et al. 2005; Tuzun, Sharp et al. 2005) and common inbred laboratory mouse strains (Li, Jiang et al. 2004; Graubert, Cahan et al. 2007; She, Cheng et al. 2008; Watkins-Chow and Pavan 2008; Yalcin, Wong et al. 2011). Although, on a genomic level, CNVs are relatively abundant in both human and mice, individual CNVs do not occur at a high frequency in the population. This infers that although CNVs are common, a particular CNV probably occurs at low frequency in the population (Itsara, Cooper et al. 2009). However, since some regions of the genome appear to be hotspots for *de novo* CNVs (e.g. repetitive regions such as those close to centromeres), it is likely that recurrent CNVs in specific locations do occur. These reoccurring CNVs could be of interest in terms of the development of disease and the maintenance of deleterious conditions in the population.

There are three primary mechanisms identified for the creation of CNVs: i) non-allelic homologous recombination (NAHR), ii) non-homologous end joining (NHEJ) and iii) fork stalling and template switching (FoSTeS). NAHR occurs during chromosome alignment along the metaphase plate during meiosis. This can cause loss or gain of segments of the chromosome in the gametes as well as inversions (*Figure 4.1*). It can also occur in mitosis leading to a mosaic population of somatic cells, a pattern that has been implicated in cancer (Notini, Craig et al. 2008). Duplicated sequences can lead to the misalignment of chromosomes and occurs at region-specific DNA blocks usually of 10-300kb in size with >95% homology. The human genome appears to have a larger proportion of duplicated regions than other organisms, suggesting that humans are more prone to recurrent CNVs. NAHR is postulated to be the most common way that de novo CNVs arise (Roth and Wilson 1986).

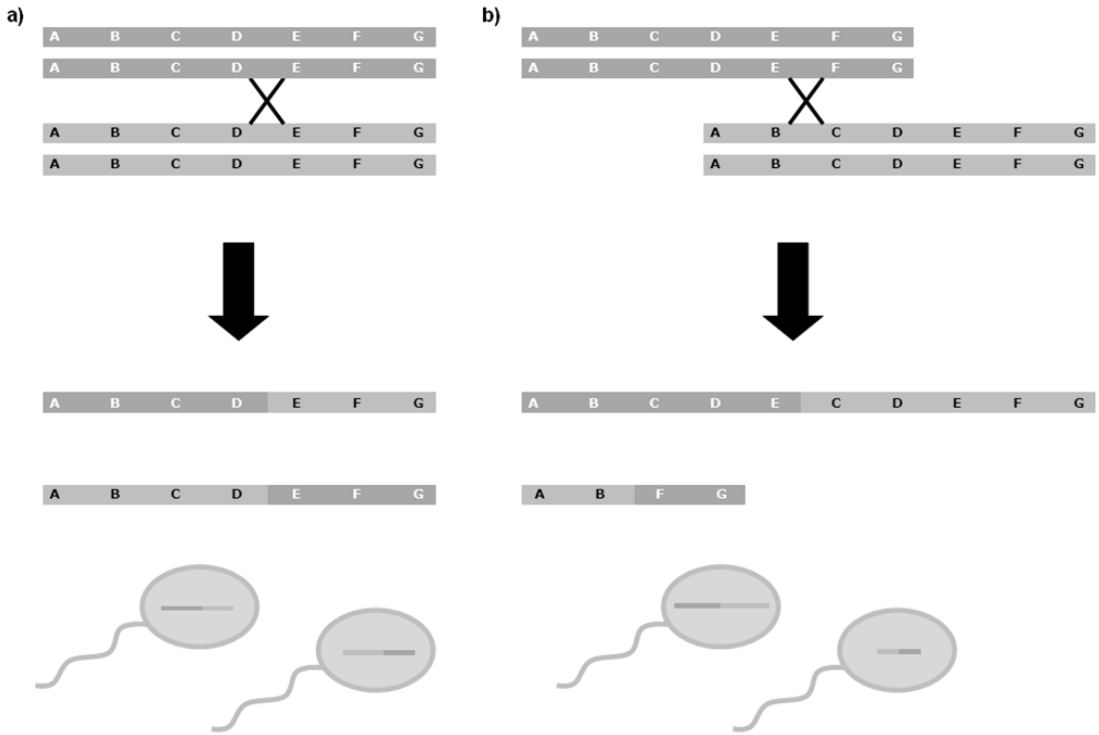


Figure 4.1 – Schematic of Recombination Events

a) Recombination occurring at the same site on sister chromatids

b) Non-Allelic Homologous Recombination at misaligned chromatids

The second mechanism implicated in the creation of CNVs is non-homologous end joining (NHEJ) (Lieber, Ma et al. 2003), which occurs when there are double strand breakage (DSB) events, and functions throughout the cell cycle act to repair such lesions (*Figure 4.2*). With no homology between stands, Ku70/80 (proteins that binds to DNA double-strand break ends and are essential for DNA repair) and the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) attach to DNA and produce blunt ends. This leaves an "information scar" at the rejoining site as ends are edited by either cleavage or the addition of nucleotides. NHEJ is a relatively frequent event due to the highly repetitive nature of the genomes of multicellular eukaryotes and because homologous recombination is too slow when sister chromatids are not aligned. NHEJ is considered to be a major mechanism in rejoining translocated chromosomes in cancer (Elliott and Jasin 2002). Combined with DSB homologous repair, NHEJ is also used to explain duplications. It is not clear how frequently DSB occurs in humans, and therefore it is unknown how common NHEJ really is, although inherited defects in NHEJ are known to account for ~15% of human severe combined immunodeficiency (Lieber, Ma et al. 2003).

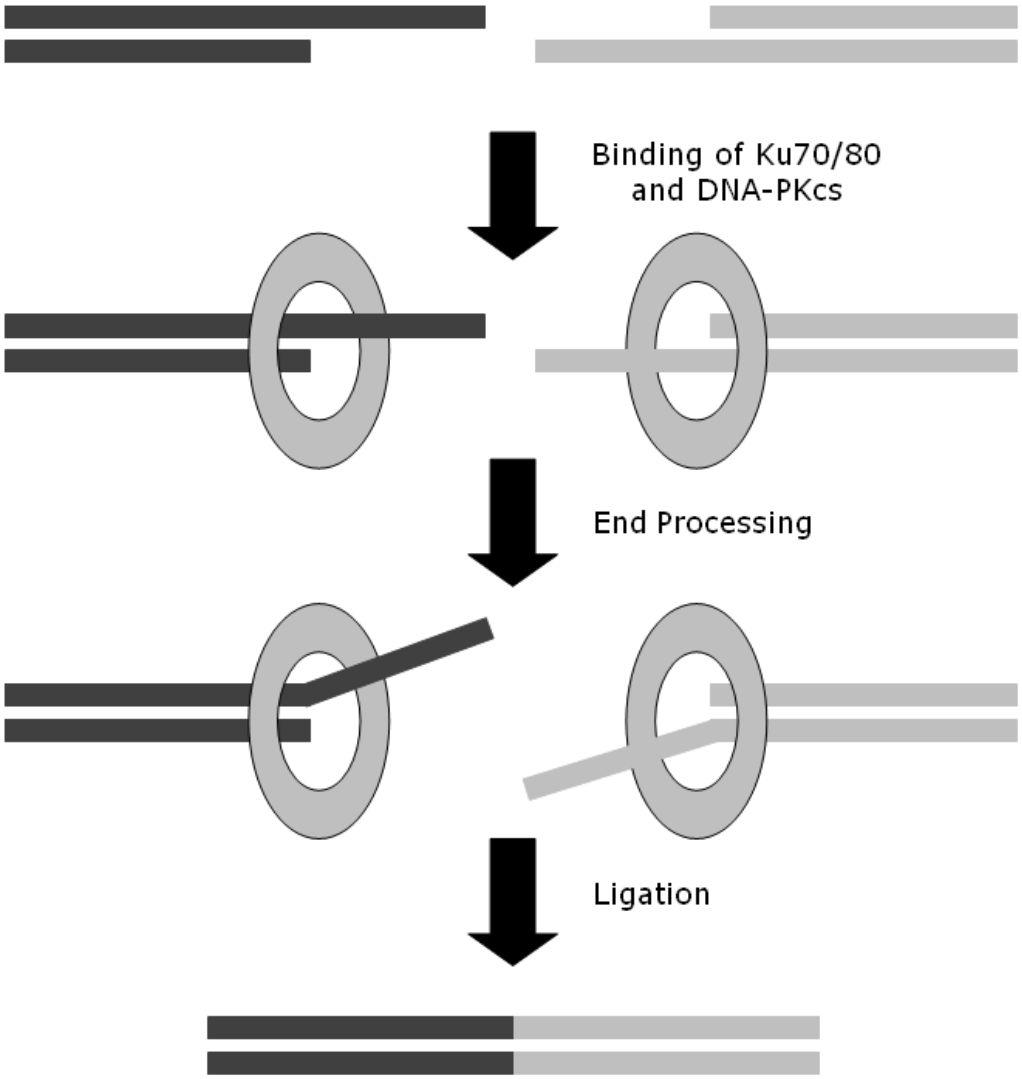


Figure 4.2 – Non-Homologous End Joining

The third mechanism in the creation of CNVs is fork stalling and template switching (FoSTeS) (Hastings, Lupski et al. 2009). This occurs during mitosis and can account for complicated rearrangements. No extensive homology is required for FoSTeS to happen, but it occurs at sites of microhomology (4-15bp). When the replication forks stall, the 3′ primer end of a DNA strand can change templates to a ssDNA template in a nearby replication fork. This causes the loss of looped out regions, resulting in one segment of the genome moving to another. This mechanism can apply to both genic and single exon rearrangements and indicates a link to gene evolution and potentially the mechanism behind exon shuffling (*Figure 4.3*).

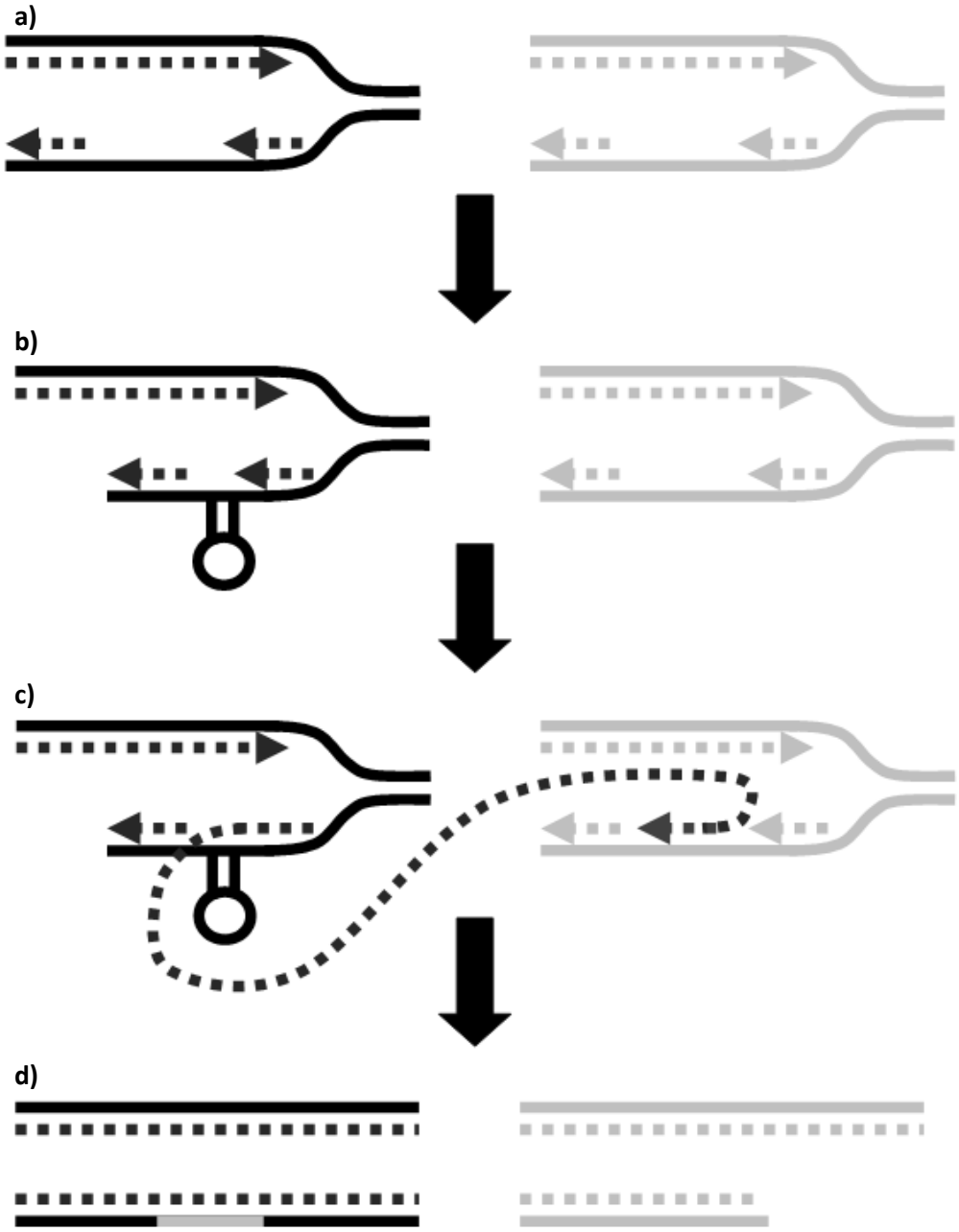


Figure 4.3 - Fork Stalling and Template Switching

a) *Replication occurring on two sections on DNA*
b) *DNA loops out on the lagging strand of one section of DNA*
c) *Microhomology on the other section of DNA causes replication of the first DNA strand to continue on to it*
d) *Results in DNA from the second section becoming part of the first and a loss of DNA in the second DNA strand*

CNVs have been shown to occur more frequently close to the telomeres and centromeres than other areas of the genome (Nguyen, Webber et al. 2006), potentially due to the repetitive nature of the DNA sequence in these regions and the reliance on repetitive sequence and regions of homology for two of the primary mechanisms underlying the creation of CNVs, NAHR and FoSTeS (Roth and Wilson 1986; Hastings, Lupski et al. 2009).

CNVs have been shown to have a strong effect on the expression of the genes they encompass. They have been found to account for 17.7% of the total detected genetic variation in the levels of gene expression in individuals from the HAPMAP project (Stranger, Forrest et al. 2007). In inter-crossed mouse studies, it was found that 83% of the genes predicted to be within CNVs were differentially expressed (Orozco, Cokus et al. 2009). According to the most recent version of the Database of Genomic Variation (http://projects/tcag.ca/variation), which defines structural variation involving segments of DNA that are >1kb in healthy individuals, there are 23,893 different CNVs at 15,931 loci throughout the human genome, suggesting that direct transcriptomic consequences resulting from these variants are likely to be widespread.

Studies in mice report more than 600 associations between CNVs and gene expression; of note, most of the CNVs causing transcriptional changes are mapped to regions outside of actual transcribed regions, indicating that CNVs do not need to directly affect the gene sequence to cause alterations in its expression (Cahan, Li et al. 2009). In fact, CNVs have been shown to alter the expression of genes up to 0.5Mb away, although genes directly in the vicinity of a genomic alteration show a greater gene expression change (Henrichsen, Vinckenbosch et al. 2009). Nearly 30% of the expression level variation between different strains of mice are hypothesised to be caused by CNVs (Cahan, Li et al. 2009). From these studies, it can be concluded that CNVs not only alter the expression of genes in the direct vicinity of the CNV, but also interfere with the expression of other nearby (and more distal) loci. This causes problems for research into the role of CNVs in phenotypic variation and disease, because it may not be necessarily clear how structural variation maps onto phenotypic variation.

4.2.2 Copy Number Variations in Disease

As expected for such a major type of genomic variation, CNVs have been shown to play an important role in the aetiology of many diseases. Several Mendelian disorders, for example, have been shown to be caused by structural genomic variations including leukodystrophy, which is caused by a duplication in the protein lamin B1 gene (Padiath, Saigoh et al. 2006), and hereditary pancreatitis, caused by a duplication in *PRSS1* (Le Marechal, Masson et al. 2006). Recently, there has been mounting interest about the role of CNVs in complex disease (WellcomeTrustSangerInstitute 2011).

Genomic alterations are a hallmark of many types of cancer and there are several rare CNVs that may be responsible for a high cancer risk (Kuiper, Ligtenberg et al. 2010). Immunodeficiency disorders are another area where CNVs have been shown to play a role. Systemic lupus erythematosus has been found to be associated with CNVs in *C4A/C4B* (Yang, Chung et al. 2007) and *CCL3L1* (Mamtani, Rovin et al. 2008), and psoriasis, rheumatoid arthritis and type 1 diabetes with a CNV in *CCL3L1* (McKinney, Merriman et al. 2008). Furthermore, increased HIV resistance is associated with CNVs in *DEFB104* (Milanese, Segat et al. 2009) and *CCL3L1* (Gonzalez, Kulkarni et al. 2005).

4.2.3 Copy Number Variation in Psychiatric Disease

Of particular relevance to this study is recent data showing that the risk of many psychiatric disorders, including those linked to advanced paternal age, are associated with structural genomic variation; both at specific sites in the genome and via an overall increased in CNV burden across the genome of affected individuals. Of note, much of the risk appears to be with *de novo* CNVs. To date, the variants associated with neuropsychiatric disease generally appear to be individually rare, but highly penetrant where they occur. Some of the evidence linking CNVs to autism, schizophrenia and BD is presented below.

Autism patients have been shown to have a higher incidence of *de novo* CNVs (~10%) than controls (~1%) and in 1st degree unaffected relatives (Sebat, Lakshmi et al. 2007). They are believed to play a role in ~7% of cases of ASD (Marshall, Noor et al. 2008), with the most

common recurrent CNV observed in autism only occurring in 1-2% of cases (Weiss, Shen et al. 2008). For autism, CNVs appear to be more associated with sporadic cases than multiplex (familial) cases (Morrow 2010). Several recurrent CNVs have been associated with autism (see *Table 4.1*), and there are examples of both inherited and *de novo* events mediating risk. Some CNVs associated with autism are not necessarily recurrent with regard to their exact location, but act to disrupt a specific gene. An example of this is with *NRXN1* where CNVs affecting the gene are significantly more common in ASD patients than the general population, but the location of the CNV differs across affected individuals (Kim, Kishikawa et al. 2008). The most widely researched region of the genome in terms of CNVs in ASD is the 15q11-13 locus. This region has been found to contain duplications in a number of studies and still remains a major candidate region for the development of ASDs, especially when maternally inherited (Cook, Lindgren et al. 1997; Cook, Courchesne et al. 1998; Sebat, Lakshmi et al. 2007). CNVs in some other regions of the genome have been identified in multiple studies of ASD, including 16p11.2 (Sebat, Lakshmi et al. 2007; Marshall, Noor et al. 2008; Weiss, Shen et al. 2008).

Table 4.1 – Examples of ASD Related Genes and Regions Containing CNVs

** Denotes where only genes affected are given instead of loci*

+ Denotes where loci affected do not give exact details of genes involved

Study	**Locus**	**Genes**
(Ghaziuddin and Burmeister 1999)	2q37+	
(Serajee, Zhong et al. 2003)	7q31	*GRM8*
(Vorstman, Staal et al. 2006)	2q37+	
	5p15+	
	11q25+	
	16q22.3+	
	17p11.2+	
	18q21.1+	
	18q23+	
	22q11.2+	
	22q13.3+	
	Xp22.2–p22.3+	

(Ullmann, Turner et al. 2007)	16p13.1^{+}	
(Bakkaloglu, O'Roak et al. 2008)	7q11.22	*AUTS2*
	7q35	*CNTNAP2*
(Kumar, KaraMohamed et al. 2008)	16p11.2^{+}	
(Fernandez, Morgan et al. 2008)	3p26.2-p26.3*	*CNTN4*
(Kim, Kishikawa et al. 2008)	2p16.3^{+}	*NRXN1*
(Morrow, Yoo et al. 2008)	4q28.3	*PCDH10*
	3q24	*NHE9*
(Weiss, Shen et al. 2008)	16p11.2^{+}	
(Doornbos, Sikkema-Raddatz et al. 2009)	15q11.2^{+}	
(Glessner, Wang et al. 2009)	2p16.3*	*NRXN1*
	3p26.2-p26.3*	*CNTN4*
	3q26.31*	*NLGN1*
	9q33.1*	*ASTN2*
	15q11.2*	*UBE3A*
	6q26*	*PARK2*
	1q25.1-q25.2*	*RFWD2*
	3q13.33*	*FBXO40*
	2q24.3*	*AK123120*
(Moreno-De-Luca, Mulle et al. 2010)	17q12^{+}	
(Roohi, Montagna et al. 2009)	chr3p26.3	*CNTN4*
(Sykes, Toma et al. 2009)	22q13.3	*SHANK3*
(Strom, Stone et al. 2010)	17q21.33*	*CACNA1G*

In schizophrenia, it has been observed that *de novo* CNVs are present in a significantly higher percentage of cases than controls (International Schizophrenia Consortium 2008; Walsh, McClellan et al. 2008; Xu, Roos et al. 2008), and the frequency is higher still in early-onset cases (Walsh, McClellan et al. 2008). In a separate study, the overall occurrence of CNVs has been shown to be 1.15 fold higher in cases than in controls (Stone, O'Donovan et al. 2008). *De novo* CNVs have been shown in some studies to be significantly associated with sporadic cases of schizophrenia, but play less of a role in familial cases (Xu, Roos et al. 2008). Other studies, however, suggest that individual rare inherited CNVs are still more common in familial cases compared to controls (Xu, Woodroffe et al. 2009). These findings suggest that

these CNVs are rare in the population as a whole but more common in schizophrenia cases either due to *de novo* events or because they have been inherited from affected parents. Some examples of genes affected by CNVs associated with schizophrenia are listed in *Table 4.2*. Again, these CNVs are individually rare; the most common CNV (22q11.2) is only seen in 0.3% of schizophrenic cases (International Schizophrenia Consortium 2008). Genes disrupted by CNVs in schizophrenia include *DISC1* and *DISC2* (Millar, Wilson-Annan et al. 2000; Blackwood, Fordyce et al. 2001; Chubb, Bradshaw et al. 2008), *PDE4B* (Millar, Pickard et al. 2005; Millar, Mackie et al. 2007), *NPAS3* (Kamnasaran, Muir et al. 2003; Pickard, Pieper et al. 2006; Pickard, Christoforou et al. 2009), *NXRN1 (International Schizophrenia Consortium 2008; Vrijenhoek, Buizer-Voskamp et al. 2008; Rujescu, Ingason et al. 2009; Ikeda, Aleksic et al. 2010)* and *CHRNA7* (International Schizophrenia Consortium 2008; Stefansson, Rujescu et al. 2008).

Table 4.2 - Schizophrenia Related Genes and Regions Containing CNVs

** Denotes where only genes affected are given instead of loci*

+ Denotes where loci affected do not give exact details of genes involved

Study	**Locus**	**Genes**
(Millar, Wilson-Annan et al. 2000)	1q42.2	*DISC1, DISC2*
(Riley, Makoff et al. 2000)	15q13-14	*CNTNAP*
		CHRFAM7A
(Wilson, Flibotte et al. 2006)	1p34.3	*GRIK3*
	5q21.3	*EFNA5*
	14q23.3	*AKAp5*
	22q12.3	*CACNG2*
(Moon, Yim et al. 2006)	3q13.12	*CD47*
	14q32.33	*IGHV3, IGHV4, IGHV7*
	15q15.1+	
	Xq23	*LOC441513*
	22q11.2+	
(Friedman, Vrijenhoek et al. 2008)	7q34-35.1+	
(International Schizophrenia	1q21.1+	

Consortium 2008)	2p16.3	*NRXN1*
	7q35	*CNTNAP2*
	12p11.23^{+}	
	15q13.1	*APBA2*
	15q13.3	*CHRNA7*
	16p12.1-12.2^{+}	
	22q11.2^{+}	
(Kirov, Gumus et al. 2008)	2p16.3^{+}	
	15q13.1^{+}	
	16p12.2^{+}	
	17q22^{+}	
(Mizuguchi, Hashimoto et al. 2008)	15p12	*HCN1, MRPS30*
	11p13	*HIPK3*
	17p12	*PMP22*
(Stefansson, Rujescu et al. 2008)	1q21.1	*GJA8*
	22q11^{+}	
	15q11.2	*CYFIP1*
	15q13.3	*CHRNA7*
(Vrijenhoek, Buizer-Voskamp et al. 2008)	2p16.3	*NRXN1*
	2p25.3	*MYT1L*
	5p15.2	*CTNND2*
	9q33.1	*ASTN2*
(Walsh, McClellan et al. 2008)	2q33.3	*ERBB4*
	3p25.1	*GRM7*
	5p13	*SLC1A3*
	7q21	*MAGI2*
	11q14.1	*DLG2*
(Xu, Roos et al. 2008)	3q22.2	*EPHB1*
	5q31.1	*RAPGEF6*
	12q24	*CIT*
	14q32.13-32.2	*DICER1*
	19q13.12	*ZNF Cluster*

	22q11.2^{+}	
(Bruce, Sachs et al. 2009)	5p15.1	*ANKH*
(Ingason, Rujescu et al. 2011)	16p13.1	*NTAN1, NDE1*
(McCarthy, Makarov et al. 2009)	16p11.2^{+}	
(Need, Ge et al. 2009)	8p22	*TUSC3, PCM1, ASAH1, NAT1, NAT2*
	15q11.2-13.3	*APBA2*
	16p13.11-12.4	*NDEL1*
	22q11.2^{+}	
(Rodriguez-Santiago, Brunet et al. 2010)	16p13.3	*SSTR5*
	22q11.23	*GSTT1, GSTT2*
(Rujescu, Ingason et al. 2009)	2p16.3	*NRXN1*
(Glessner, Reilly et al. 2010)	9q34.4*	*CACNA1B*
	10q11.21*	*RET*
	18q12.3*	*RIT2*
	16p11.2*	*DOC2A*
(Ikeda, Aleksic et al. 2010)	1q21.1^{+}	
	16p13.1^{+}	
	2p16.3*	*NRXN1*
(Mulle, Dodd et al. 2010)	3q23	*PAK2, DLG1*
(Levinson, Duan et al. 2011)	3q29^{+}	
	7q36	*VIPR2*
(Vacic, McCarthy et al. 2011)	7q36.3*	*VIPR2*

Although BD has also been associated with an increased burden of CNVs (Zhang, Cheng et al. 2009), the evidence is considerably weaker than for either ASD or schizophrenia, and is primarily in the context of early-onset forms of the disorder (Priebe, Degenhardt et al. 2011). This is interesting because BD, like autism and schizophrenia, is also associated with advanced paternal age. There is little evidence that CNVs associated with schizophrenia also confer a risk for BD; in some cases, schizophrenia-associated CNVs are actually less common in BD than in controls (Grozeva, Kirov et al. 2010; McQuillin, Bass et al. 2011) although in Grozeva *et al's* study, the number of CNVs in BD was not increased compared with controls and was

significantly less than in schizophrenia cases (Grozeva, Kirov et al. 2010), and in McQuillin *et al's* study the rate of CNVs in BD cases was lower than in controls (McQuillin, Bass et al. 2011). Some examples of genes affected by CNVs associated with BD are listed in *Table 4.3*.

Table 4.3 – Bipolar Disorder Related Genes and Regions Containing CNVs

Study	**Loci**	**Genes**
(Wilson, Flibotte et al. 2006)	1p34.3	*GRIK3*
	14q23.3	*AKAp5*
	22q12.3	*CACNG2*
(Lachman, Pedrosa et al. 2007)	3q13.3	*GSK3β*

From the studies discussed above and others, there is overall strong evidence to support the role of CNVs in ASD, schizophrenia and BD. All three disorders show evidence for the role of recurrent CNVs, rare CNVs, *de novo* CNVs and overall load of CNVs in their aetiology. Many studies show consistent or overlapping regions containing CNVs, whereas others show 'exclusive' CNVs that are not found in other studies. CNVs at some loci are recurrent across multiple disorders such as 1q21.1, 15q13.3, 16p11.2 and 22q11.2, which have been associated with both schizophrenia and autism (Weiss, Shen et al. 2008), and 1p34.3, 14q23.3 and 22q12.3, which have been seen in both schizophrenia and BD (Wilson, Flibotte et al. 2006). This suggests that in some instances these disorders may share common aetiological pathways.

If CNVs are responsible for the development of psychiatric disease, the mechanism behind this link has to still be elucidated. Either CNVs could lead to a general impairment in cognitive function as an intermediate phenotype, ultimately resulting in disorders such as autism and schizophrenia, or specific CNVs could be directly responsible for the disorder itself either individually or as a pleiotropic factor (O'Donovan, Kirov et al. 2008). Interestingly, brain expressed genes appear to be more protected from CNV induced gene expression changes than genes expressed in other tissues (Henrichsen, Vinckenbosch et al. 2009). The parent-of-origin of CNVs may be of importance in the development of disease. For example, in a recent CNV screen of patients with intellectual disability, 76% of the phenotype-associated *de novo* CNVs were on the paternally inherited allele, especially CNVs over 1Mb and deletions (Hehir-

Kwa, Rodriguez-Santiago et al. 2011), indicating that CNVs inherited from the father could be a larger contributor to the development of psychiatric disease. This observation could have implications for understanding the mechanism(s) involved in mediating the paternal age effect in neuropsychiatric disorders.

4.2.4 Copy Number Variations and Paternal Age

As male germline cells undergo significantly more divisions than female germline cells (section 1.4), the potential for more *de novo* mutations coming from the father is greater than from the mother (Crow 1999). This includes both point mutations and CNVs. The incidence of *de novo* CNVs in the offspring has been previously hypothesised to increase with advancing paternal age (Walter, Intano et al. 1998), and evidence from evolutionary sequence data indicates that the overall rate of deleterious mutation observed in older fathers may be high enough to have a large effect on human well-being (Crow 2000).

The link between advanced paternal age and structural genomic variation has not been widely explored. Paternal age has been shown to be associated with a higher incidence of balanced translocations occurring in the sperm (Thomas, Morris et al. 2010), although this observation has not been ubiquitously reported (Thomas, Durkie et al. 2006). In mice, older males have been shown to have an increased incidence of sperm mutations than younger mice (Walter, Intano et al. 1998). To date, however, only one association has been reported between the offspring of older fathers and the offspring of younger fathers in humans, finding that increased paternal age was only associated with an increased burden of CNVs not flanked by segmental duplications (Hehir-Kwa, Rodriguez-Santiago et al. 2011). Only one study in mice has been previously undertaken. In this study, seven *de novo* CNVs were observed in six offspring of fathers 12-18 months old and no *de novo* CNVs in the offspring of fathers three months old (Flatscher-Bader, Foldi et al. 2011). The paucity of research into this area highlights the need for further studies into the link between advanced paternal age, increased CNV burden, and risk of neuropsychiatric disease.

4.3 Aims

By systematically investigating CNVs across the genome in the offspring of mice of two different ages (two months and ten months), the aims of this chapter were to i) investigate whether paternal age is associated with an increased incidence of CNVs in the offspring and ii) whether there are any recurrent CNVs occurring in the context of advanced paternal age. Where any *de novo* CNVs are identified, I aimed to assess their relevance to the paternal-age associated behavioural changes reported in Chapter 3 via functional pathway analyses.

4.4 Methods

Several methods have been developed for the detection of CNVs across the genome. The earliest forms of detection employed cytogenetic approaches and fluorescence *in situ* hybridization (FISH), which uses fluorescent probes that bind to specific regions of the genome that can be visualised using fluorescence microscopy. There are limitations to these approaches; FISH can only be used to visualise large structural changes to the genome and it does not have the resolution to identify smaller CNV events such as those commonly associated with complex disease phenotypes. Furthermore, the target regions have to be already known and so this approach cannot be used to screen the genome for *de novo* events. Microarray-based methods such as CGH and high-resolution SNP arrays have revolutionised CNV analysis as they allow detection of smaller events across multiple individuals and can be either targeted towards regions of interest or used to screen the entire genome. CGH arrays compare hybridisation to a high-resolution oligo microarray between a sample versus a common reference which are labelled by different dyes (in this study, Cy-3 and Cy-5). However, they are subject to background noise problems and are potentially subject to a high number of false positives (and negatives). The recent development of next generation sequencing approaches will enable the characterisation of the genome for CNVs at a much higher resolution, allowing the better identification of break points with less false positive results. However, this method is still very expensive and not feasible for large numbers of samples. Specific CNVs can also be assessed using quantitative PCR (qPCR), but this approach is less useful in the analysis of complex phenotypes because individual CNVs are rare and are ideally identified using a genome-wide screen. For this study we decided that the optimal strategy was to use array-CGH, employing high-resolution microarrays spanning the mouse genome.

4.4.1 Array Specifications

The arrays used in this study were Nimblegen Mouse CGH 3x720K Whole-Genome Tiling Arrays. These arrays contains over 720,000 probes covering the mouse genome based on the UCSC mm9 mouse genome build and can run three samples in parallel on a single array. In this array the probe length is 50-75mer and median probe spacing of 3537bp.

4.4.2 DNA Extraction

Spleen DNA was extracted using the Qiagen DNeasy extraction method as outlined in section 2.2.1. All DNA samples underwent stringent quality control assessment using gel electrophoresis (section 2.5) and Nanodrop spectrometry before being used in the subsequent array-CGH experiments. All included samples were of high structural fidelity and showed no evidence of degradation.

4.4.3 DNA Labelling

DNA was labelled using a protocol adapted from the Roche Nimblegen Dual Colour labelling kit (*Figure 4.4*) (Nimblegen 2011). All reagents were thawed on ice and Cy3- and Cy5-Random Nonamers were kept protected from light at all time. To prepare random nonamers, 1.75µl β-Mercaptoethanol was added to 998.25µl of Random Prime Buffer in a fume hood then 462µl of the Random Primer Buffer with β-Mercaptoethanol was added to Cy3-Random Nonamers and Cy5-Random Nonamers. The resuspended Nonamers were then made into 40µl aliquots in small PCR tubes and stored at -20^{o}C until needed. 1.5µg of the test sample DNA was added into 40µl Cy3-nonamer PCR tube and the reaction volume made up to 80µl with DNase free water. 1.5µg of the common reference sample DNA (an unused female breeder) was added into 40µl Cy5-nonamer PCR tube and the reaction volume made up to 80µl with DNase free water. The samples were heat-denatured in a thermocycler at 98°C for 10 minutes and then immediately placed in an ice water bath for 2 minutes. The dNTP/Klenow master mix was prepared and amounts per sample were as below:

dNTP Mix (10 mM each dNTP)	10µl
Klenow Fragment (3'->5' exo-) 50U/µl	2µl
Water	8µl
Total	20µl

20µl of the dNTP/Klenow master mix was added into each PCR tube and pipette–mixed before being incubated for 2 hours at 37°C in the thermocycler with the heated lid. After removing the samples from the thermocycler, 10µl of Stop Solution was added to each tube followed by the addition of 11.5µl of 5M Sodium Chloride (NaCl). All 121.5µl of sample was then added to 1.5ml eppendorf tube containing 110µl of 100% isopropanol which was then vortexed and incubated for 10 minutes at room temperature protected from light. After incubation the tubes were spun in a centrifuge at 13000rpm for 10 minutes. The supernatant was carefully removed from the tube with a pipette and discarded before the pellet was rinsed with 500µl of 80% ice cold ethanol then centrifuged at 13000rpm for 2 minutes. The supernatant was removed with a pipette and discarded. The pellet was then dried in a SpeedVacuum at 2000rpm at 65^{o}C.

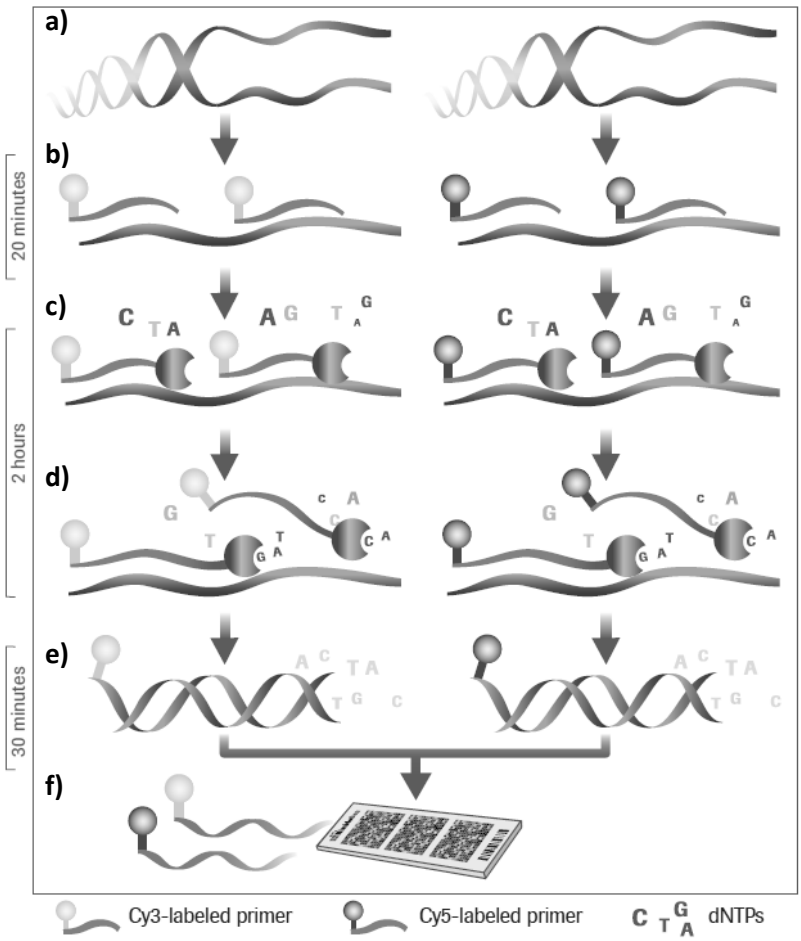


Figure 4.4 – An Outline of the DNA Labelling Protocol for Array-CGH

(Nimblegen 2011)

a) Begin labelling protocol with double-stranded test and reference DNA

b) Denature double-stranded DNA and anneal Cy3- and Cy5- random primers (9mers)

c) Synthesis begins as the Klenow fragment extends from random primers

d) Primer extension, strand displacement, and amplification

e) Remove free nucleotides (dNTPs) and other components

f) Test and reference samples were combined, denatured and hybridized to the array

4.4.4 Array Hybridization

After the test and common reference samples were labelled they were prepared for hybridization to the microarray. The pellets were rehydrated with 25µl of DNase free water and allowed to dissolve fully then vortexed and spun down to collect contents at the bottom of the tube. The concentration of each sample was measured using the NanoDrop NT-1000 using analysis DNA-50 to confirm there was >31µg of DNA for each sample required for the

arrays. 31µg of the test sample and common reference sample were combined in a 1.5ml tube and the sample mix then dried in the SpeedVacuum at 2000rpm 65°C. The sample was then rehydrated using 5.6µl of DNase free water, vortexed and spun down to collect at the bottom of the tube. The hybridization Solution Master Mix was made per array as below.

2X Hybridization Buffer	35µl
Hybridization Component A	14µl
Alignment Oligo	1.4µl

14.4µl of Hybridization Solution Master Mix was added to each sample tube and vortexed for 15 seconds then spun down. The tubes containing the sample mix were then incubated in a heat block at 95°C for 5 minutes and placed in the Hybridization System preheated to 42°C. The HX3 Mixer was assembled on to the array using the Precision Mixer Alignment Tool (PMAT) (*Figure 4.5*). The mixer was sealed using the braying tool on the 42°C hybridisation station.

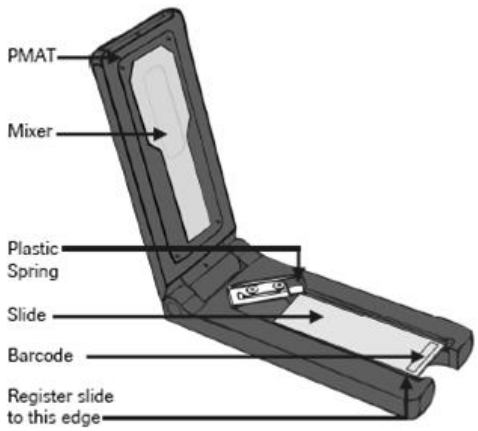


Figure 4.5 - PMAT with Mixer and Array

To load the sample, the array was placed into a hybridization chamber and using a positive displacement pipette, 18µl of Hybridization mixture was loaded into the array through each of the HX3 mixer fill ports (*Figure 4.6*). The array was then hybridised for 72hrs at 42°C.

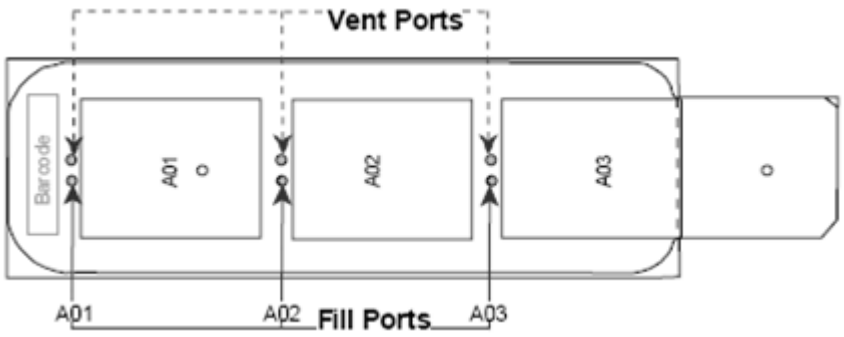


Figure 4.6 - HX3 Mixer and Array Assembly

4.4.5 Washing

After the arrays were hybridised they were washed to remove unhybridised sample. Wash buffers were prepared as below.

Mixer Removal Wash

Water	225ml
10X Wash Buffer I	25ml
1M DTT	25ul

Slide Washing

Water	22.5ml
10X Wash Buffer I, II or III	2.5ml
1M DTT	2.5ul

The Mixer removal wash was heated to 42^{o}C and poured into the removal bath. The assembled array with mixer was removed from the Hybridization System and put into the Disassemble Array Tool (DAT) and then placed together into 42°C Wash I to remove the mixer from the array (*Figure 4.7*). The slide was removed from the DAT and placed into the slide rack in room temperature Wash I and the 2 minute wash was begun by shaking array washer vigorously. The arrays was removed from Wash I and placed in room temperature Wash II and vigorously washed for 1 minute. The array was removed from Wash II and placed in room

temperature Wash III and washed for 15 seconds. The array was then spun dry using an Array-It slide dryer for 30 seconds.

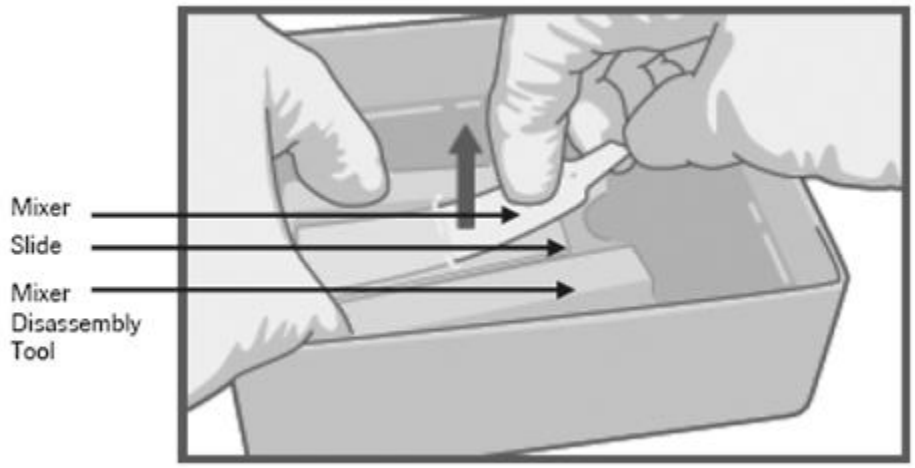


Figure 4.7 - Using Disassemble Array Tool

4.4.6 Scanning

The array was loaded into a dual laser (532nm and 635nm) GenePix 4000B scanner face down and the scanning parameters controlled using the GenePix software. The array was preview scanned and the laser voltage parameters adjusted to optimise the array image. The array was then fully scanned and both images saved for quality control and analysis.

4.4.7 Nimblescan Analysis and Data Cleaning

Data pre-processing steps were carried out using the NimbleScan Software v2.6. As the microarrays used in my experiments were 3x720k probe arrays, data from three samples was contained in each image file. The first step was to 'burst' the images so only one sample was analysed at a time. To burst images, files to be burst were selected in the Burst Multiplex Image option, and opened with the Multiplex annotation file provided with the arrays. Once burst, images for both colour channels (532nm and 635nm) were aligned and checked so that raw intensity data and probe signal data could be obtained (*Figure 4.8*). The data were then analysed using the Nimblescan CGH-segMNT algorithm to identify copy number changes by comparing the $\log_2$ ratio data of the two colour channels.

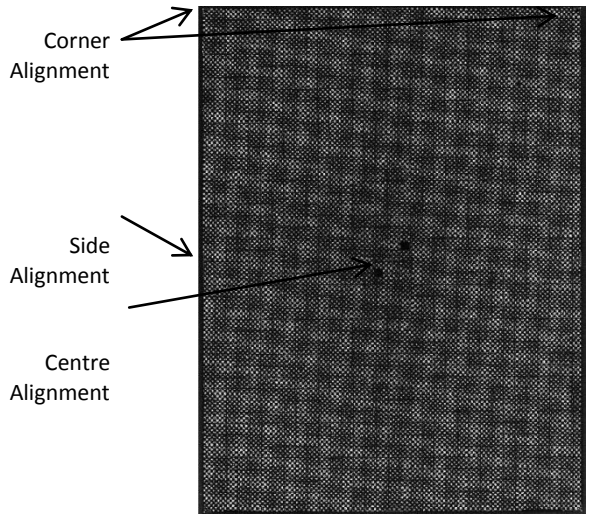


Figure 4.8 – Nimblescan Array Alignment Showing an Example Scan

As an initial quality control step, array results were initially checked to ensure they showed the correct profile across probes on the sex chromosomes. As the common reference sample was female, male samples were expected to show a clear deletion spanning the majority of probes on the X chromosome (except for the pseudo-autosomal region) and excess signal across the Y chromosome (*Figure 4.9*). All samples showed the correct dosage of sex chromosomes.

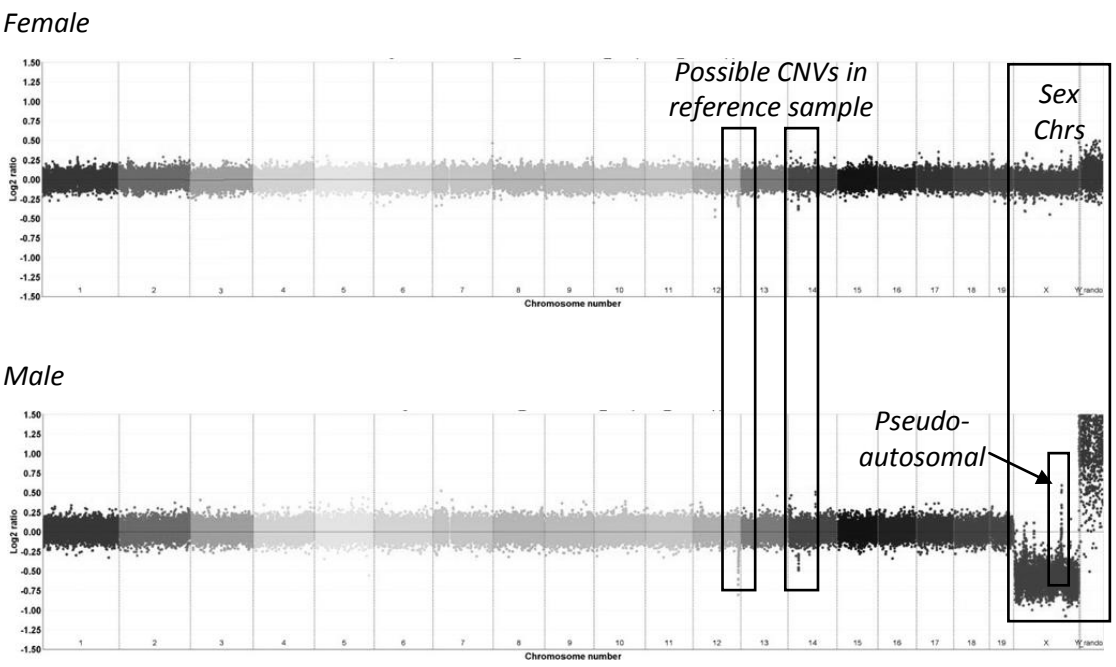


Figure 4.9 - Example Log2 Graphs of Female and Male

Since the common reference sample used in my array-CGH experiments was female, a female hybridised against the reference will show no copy number change across the X-chromosome, with a noisy signal pattern across the Y-chromosome. A male sample shows an apparent deletion across the X-chromosome and a noisy duplication across the Y-chromosome. Examples of possible CNVs in the reference sample can also be seen in this figure.

4.4.8 Further Microarray Quality Control

For further quality control (QC), raw intensity values were log2 transformed and plotted to identify samples with large variations in intensity or extremes of intensity values (*Figure 4.10*). Since all arrays are run against a common reference, the analysis involves no normalisation across arrays.

Figure 4.10 – Boxplot of Intensities from all Arrays

Log2 transformed intensity values for all samples (s) and reference (r) channels on arrays. Black line in the centre of box represents the median value

Samples were plotted against their reference to assess the overall correlation between colour channels, both with sex chromosomes (*Figure 4.11*a) and without (*Figure 4.11*b). This was to check for sex differences where applicable, because, as described, the common reference for all samples was female. As the reference sample was the same across all arrays, the log2 intensities were plotted against a single reference array's signal (PA52) to assess array quality. The plots were correlated across the vast majority of samples, and any arrays showing a poor correlation were removed from further analysis (*Figure 4.12*).

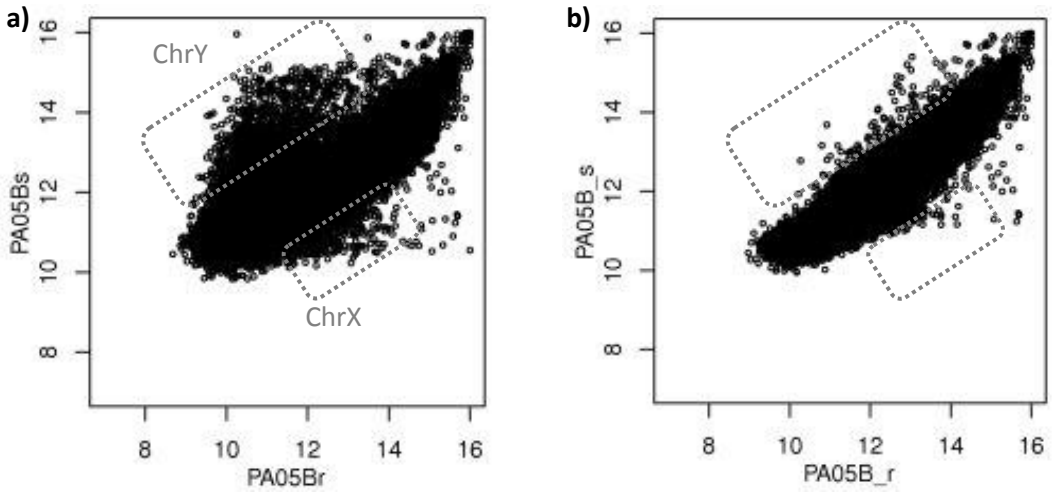


Figure 4.11 – Correlation of Log2 Intensities - Male Samples vs. Female Reference

Reference (_r) and sample (_s) plotted to show their correlations

a) *With sex chromosomes – High correlation but many outliers which are probes on the sex chromosomes as reference is female and sample is male.*
b) *Without sex chromosomes.*

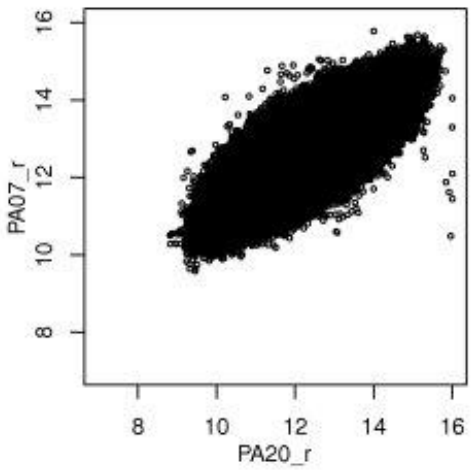


Figure 4.12 – Plot of Log2 Intensities of Reference versus Reference

All reference log2 intensities were plotted against the reference from PA20 (x-axis) to check for correlations. As the reference for all arrays was the same female sample, probe intensities should be highly correlated with any variance due to technical/experimental factors.

4.4.9 Analysis

Data cleaning and analysis was carried out using the Partek® Genomics Suite™ (2008). First the data was subjected to GC wave correction to remove any methodological based "waviness" of the array data (*Figure 4.13*), allowing for more precise calling of peaks in the CNV data (Diskin, Li et al. 2008). Waviness is a common problem in CGH and SNP genotyping array CNV experiments (Marioni, Thorne et al. 2007). Once the data was smoothed a segmentation analysis was applied to compare each probe to adjacent probes and identify regions of copy-number change.

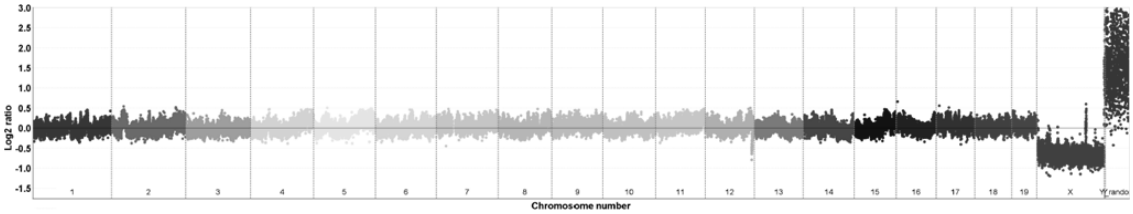


Figure 4.13 – Example of a 'Wavy' Array, Pre-Correction

Wave artefacts were observed in some arrays. These samples were wave corrected using a function within the Partek software.

De novo CNVs were identified by excluding CNVs that were clearly present in (and thus presumably inherited from) either parent. Genes affected by CNVs were annotated using Galaxy (Blankenberg, Von Kuster et al. 2010; Goecks, Nekrutenko et al. 2010). Galaxy is an open, web-based platform for data intensive biomedical research. It is linked to UCSC genome browser and in this instance was used to find genes intersecting the chromosomal positions of observed CNVs. Chromosomal positions were loaded into Galaxy using the UCSC genome browser table browser (Kent, Sugnet et al. 2002) and intersected with the entire mouse genome (mm9). This output was then combined back with the entire *de novo* CNV list.

Due to the possibility of probe location error (e.g. resulting from repetitive regions), any CNV that had an overlap with a CNV in either parent was classed as inherited and removed from the *de novo* CNV analysis as shown in *Figure 4.14*. Due to the highly repetitive nature of the Y chromosome (and the lack of Y-chromosome DNA in the female common reference sample), I chose to remove Y-linked probes from all analyses. Furthermore, due to the combination of females and males compared against a female reference, the X-chromosome was also removed from analysis, but retained as a QC measure to check sex (as described above).

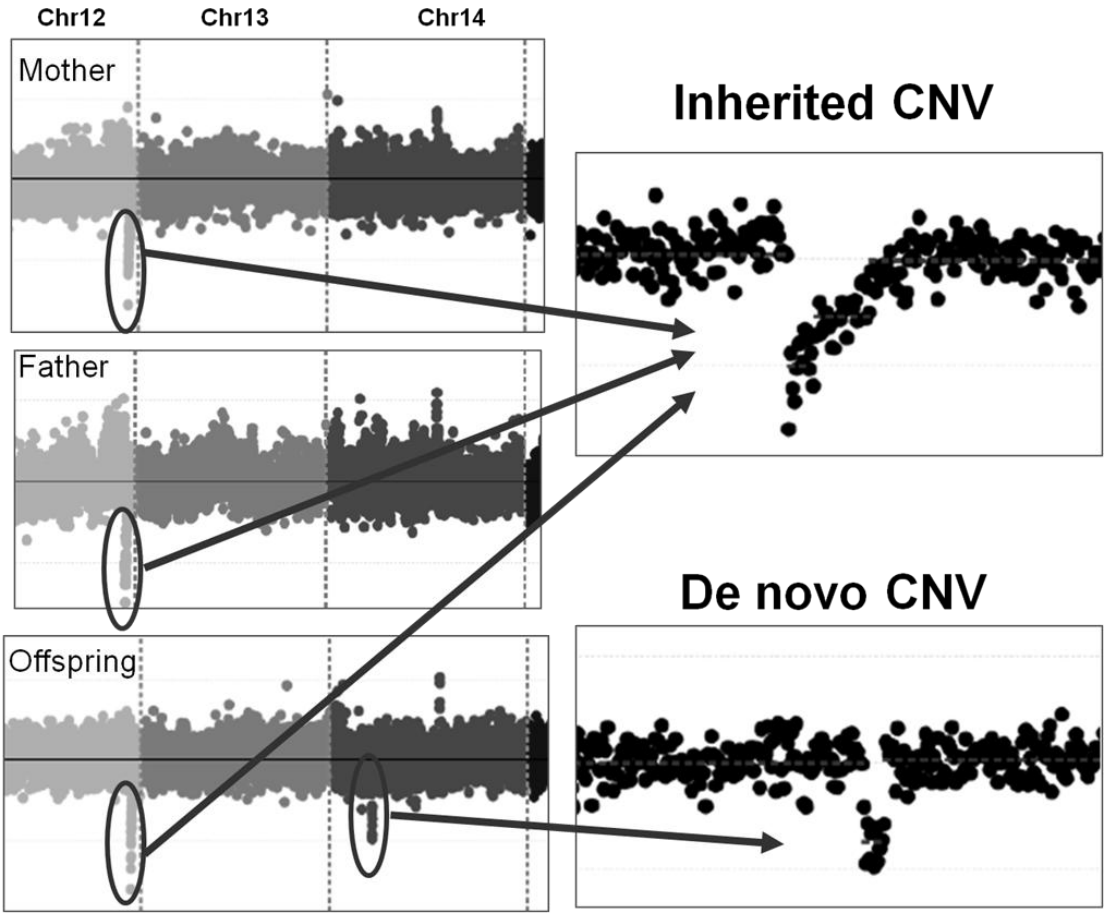


Figure 4.14 - Example of Inherited and De Novo CNV in a Trio of Mice

Traces of an offspring compared against their parents to identify inherited and de novo CNVs.

4.5 Results

4.5.1 Data Description

The first result of note is the overall high number of CNVs observed in both the breeders and offspring. As the mice were an inbred strain, obtained from the same breeding stock, the number of CNVs observed between individuals was expected to be relatively low. Regardless of paternal age, CNVs were seen in most individuals, and in most cases we observe one or more CNV. In 60 samples we see the incidence of 780 CNVs, 93 in female breeders, 292 in male breeders, and 395 in all offspring, 383 of which were non-overlapping CNVs. Taken together, we observed an overall higher average number of CNVs in the male breeders (*Figure*

4.15), where the average number of CNVs per individual was 29, compared to the offspring and female breeders where the average number of CNVs was 10 ($F(2, 58) = 5.07$, $p = 0.0094$). Since my experimental design meant that the samples were randomised across microarrays, it is unlikely that this elevated number is due to a technical artefact. Furthermore, none of the male breeder samples appeared to be experimental outliers in stringent QC measures described in section 4.4.8.

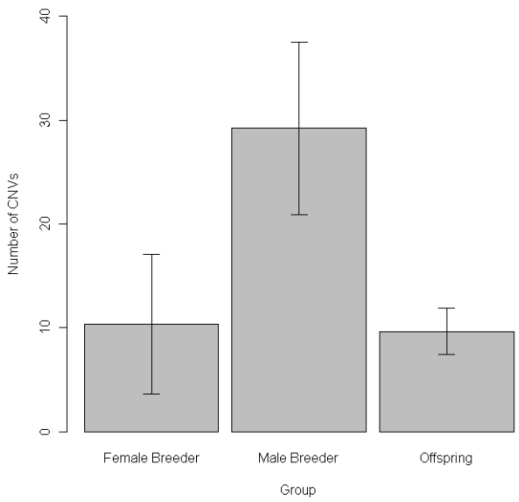


Figure 4.15 – Number of CNVs by Breeder or Offspring

The total number of CNVs for each individual sample is shown in *Figure 4.16*. Here it can be seen that several samples are clear outliers for their total burden of CNVs, including PA02 (*Figure 4.17*) (91), which reached 3 SDs above the mean and PA16B (66), PA23B (64), PA04B (60) and PA06B (68), which reached 2 SDs above the mean. Even though these arrays are of good quality based on QC measures, these samples (predominantly old male breeders) were removed from the CNV analyses presented below.

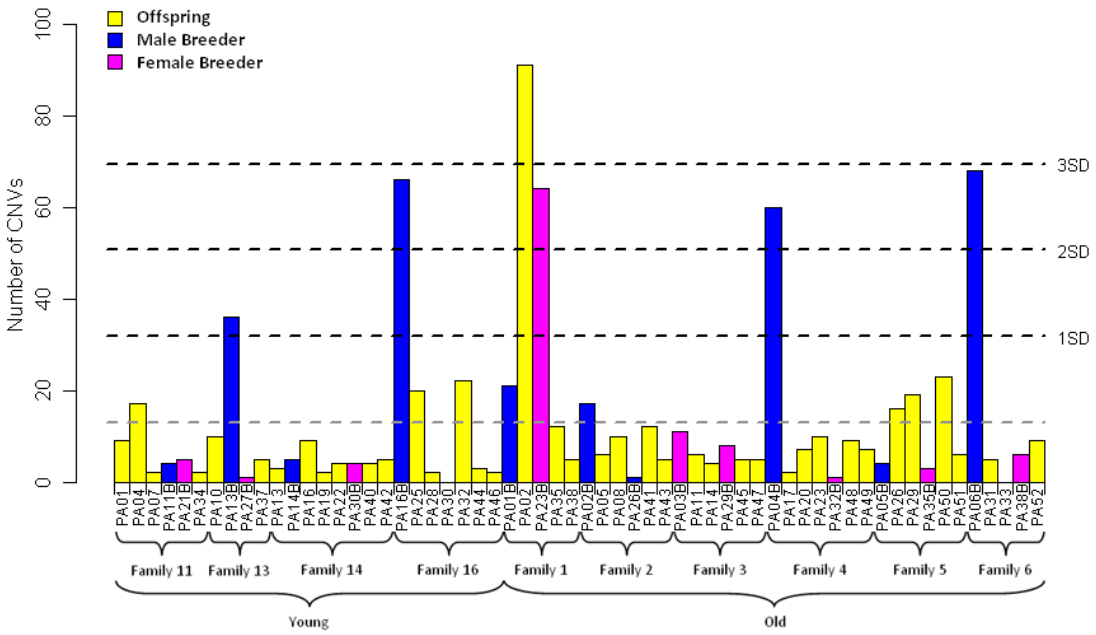


Figure 4.16 - All CNVs by Family and Group

Plot of all CNVs in all individual samples. Grey dotted line indicates mean, black dotted lines indicates standard deviations above the mean.

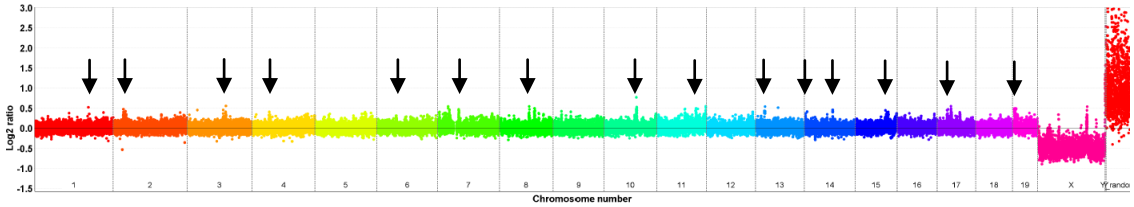


Figure 4.17 – Plot of Log2 Ratio of Sample over Reference

Plot of PA02 which has a large number of CNVs with some examples of CNVs indicated by arrows. The array is of good quality: signal intensities are high, there is no 'waviness' and the sex chromosomes show the correct dosage.

The CNV abnormalities in certain male breeders do not appear to be manifest in the germline of these individuals, as the high burden is not seen in their offspring, suggesting a somatic event potentially restricted to the spleen. Of note, the high number of CNVs could be caused by lymphomas in the spleen which are known to occur sporadically in mice and

especially with advanced age (Murphy 1968). In fact, a major cause of natural death in C57BL/6 mice is lymphoma (Brayton, Treuting et al. 2012). If these CNVs in the breeders were also seen in other tissues, especially the sperm, we would expect a higher number to be inherited by the offspring, and in all cases except PA02, the number of CNVs is significantly lower in their offspring. Once these samples were removed, we see an average of 8 (+/- 7) CNVs per individual. It is worth noting that this estimate is likely to be slightly inflated due to CNVs occurring in the reference sample (as exemplified in *Figure 4.9*); however since I removed overlapping CNVs these should each only be counted once in assessments across all samples.

Other studies looking at CNVs in populations of 'genetically identical' inbred mice have also observed a notable CNV burden. In one study using a lower resolution CGH array (385k), for example, 38 CNVs were observed between 14 colonies of C57BL/6 inbred mice (Egan, Sridhar et al. 2007). While the majority of these CNVs were rare, private events, some are common – for example a CNV in the *Ide* gene has been shown to be present in 64% of mice assayed (Watkins-Chow and Pavan 2008). Compared to other studies in inbred mouse strains, however, we see an overall higher rate of CNV burden across the genome, although this may be a result of the higher-resolution arrays used in our study compared to previous published reports.

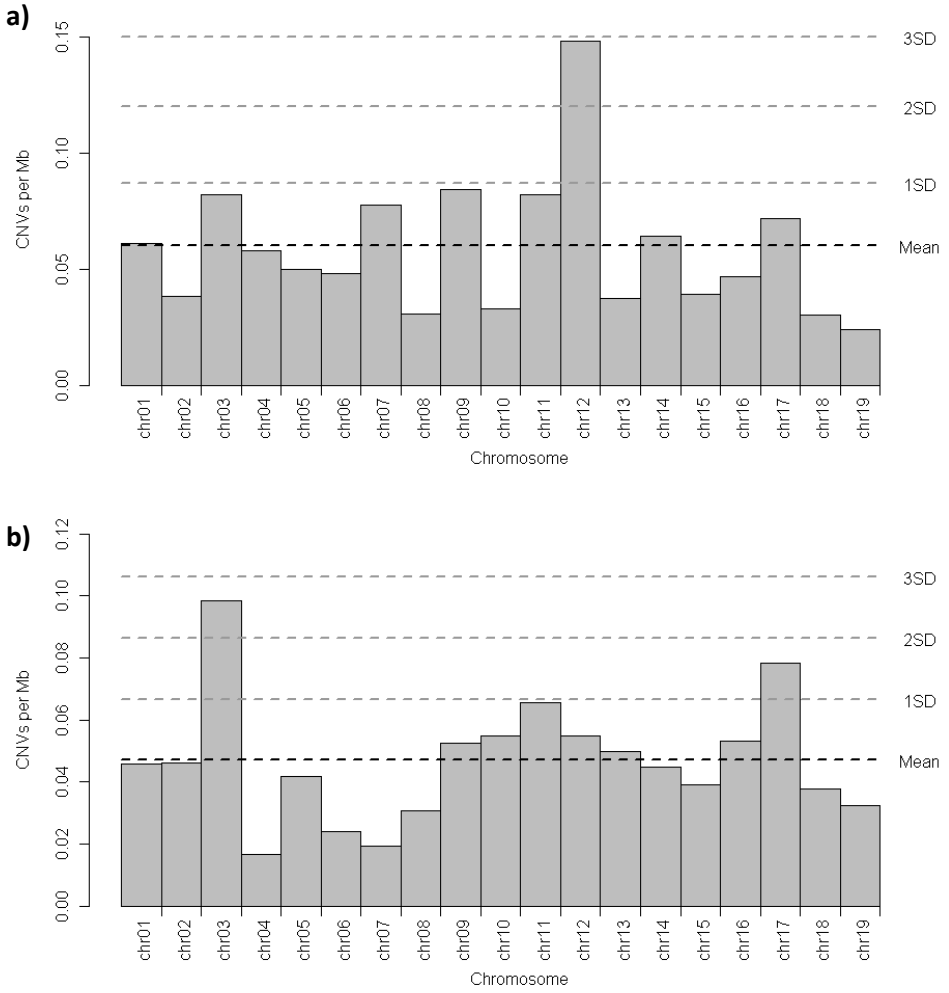
Some CNVs appear to be highly recurrent between individuals suggesting these may actually represent a CNV in the common reference sample. Some examples of this are shown in the male of *Figure 4.9*. As these CNVs occurred in at least one parent per family, they were removed from analysis of *de novo* CNVs. To investigate whether some chromosomes had an abundance of CNVs across all individuals, all non-overlapping CNVs were separated by chromosome. The mean number of non-overlapping CNVs per Mb was 0.06. No chromosome reached three SDs above the mean but chromosome 12 had a number of CNVs more than two SDs above the mean (*Figure 4.18*) suggesting it may be more prone to the occurrence of CNVs. Looking at non-overlapping *de novo* CNVs (i.e. those present in the offspring but not parents in informative families) by chromosome, the mean number of CNVs per Mb was 0.05, with the highest rate observed on chromosome 3.

Figure 4.18 - Number of Non-Overlapping CNVs by Chromosome in all Animals

Plot of non-overlapping CNVs per Mb plotted by chromosome with outliers removed. Black dotted line indicates mean, grey dotted lines indicates standard deviations above the mean

a) All non-overlapping CNVs

b) All de novo non-overlapping CNVs



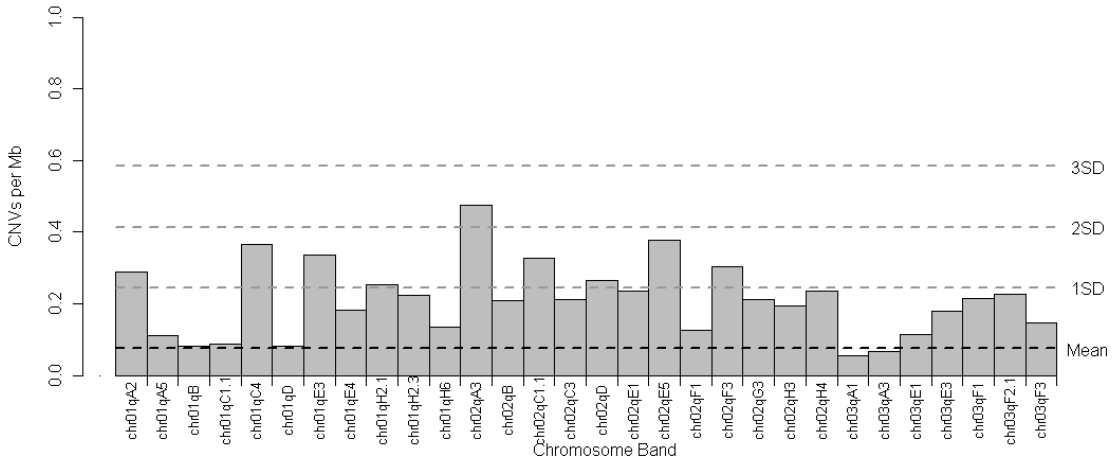
To look for regions of the genome that are enriched for CNVs, I also divided the CGH data by discrete chromosomal bands. The mean number of all non-overlapping CNVs per Mb including all chromosomal bands in the genome was 0.08, so to investigate whether certain chromosome bands contained more CNVs, a value of two SDs above the mean was used as an arbitrary threshold (0.41 per Mb); these were classed as “CNV rich”. Eight bands reached the mean plus 2SD threshold (2qA3, 4qC2 7qF5, 9qA5.2, 12qA1.3, 15qB2, 15qD2 and 15qD3, 17qB1), with three bands reaching a value of 3SD above the mean (11qB4, 18qD2 and 19qC2).

This implies that these regions may be more prone to CNVs as they occur more commonly in these bands than the rest of the genome. These regions are shown in *Figure 4.19*.

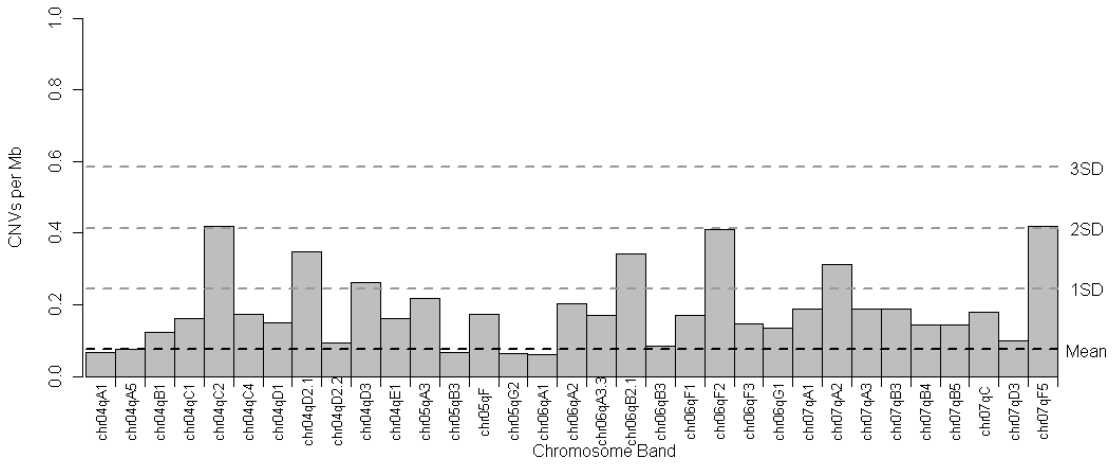
Figure 4.19 - CNVs by Chromosome Band

Plot of non-overlapping CNVs per Mb plotted by chromosome band. Black dotted line indicates mean, grey dotted lines indicates standard deviations above the mean.

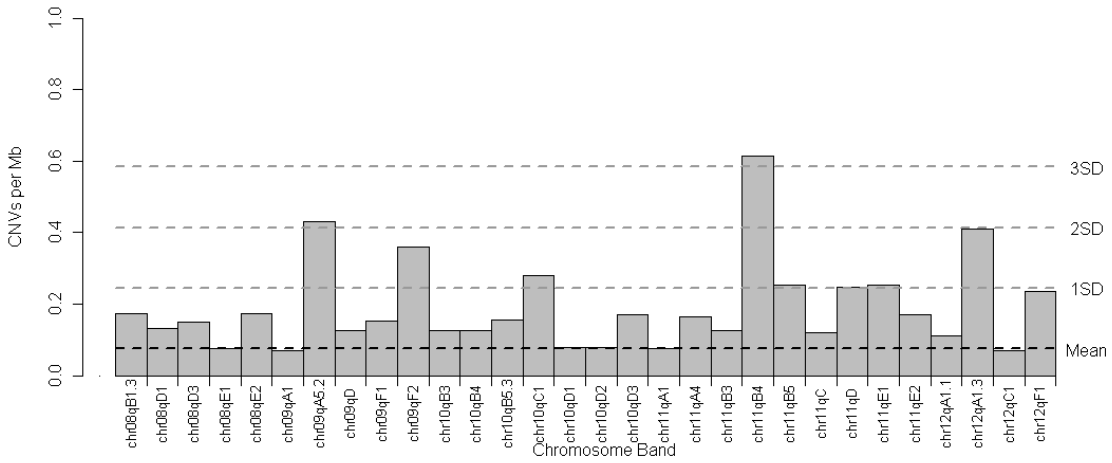
Chromosomes 1-3



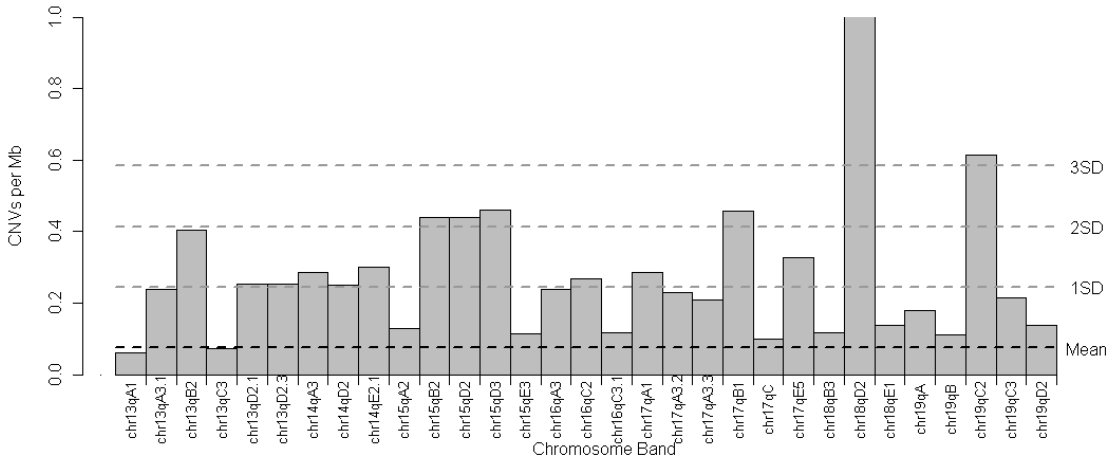
Chromosomes 4-7



Chromosomes 8-12



Chromosomes 13-19

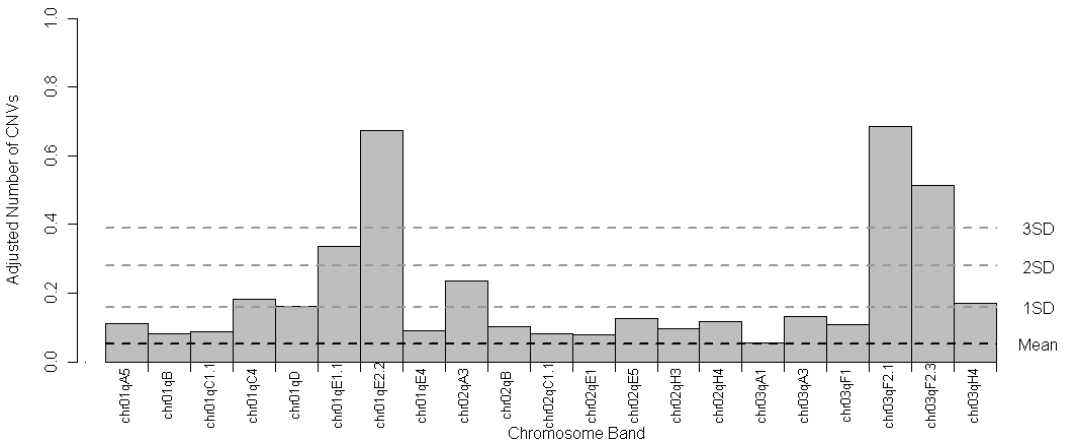


To identify chromosome bands with a higher frequency of *de novo* CNVs, all non-overlapping *de novo* CNVs were split by band and divided by band size in Mb to give CNVs per MB by band. The mean value of CNVs per Mb including all chromosomal bands in the genome was 0.05. 1qE2.2, 3qF2.1, 3qF2.3, 6qF2, 9qA5.3, 15qF3, 17qD and 18qB2 all exceeded a level three SDs above the mean (0.39 CNVs per Mb) and 16 bands reached a value 2SDs above the mean (0.28 CNVs per Mb) (*Figure 4.20*), suggesting that these regions may be more prone to the occurrence of *de novo* CNVs.

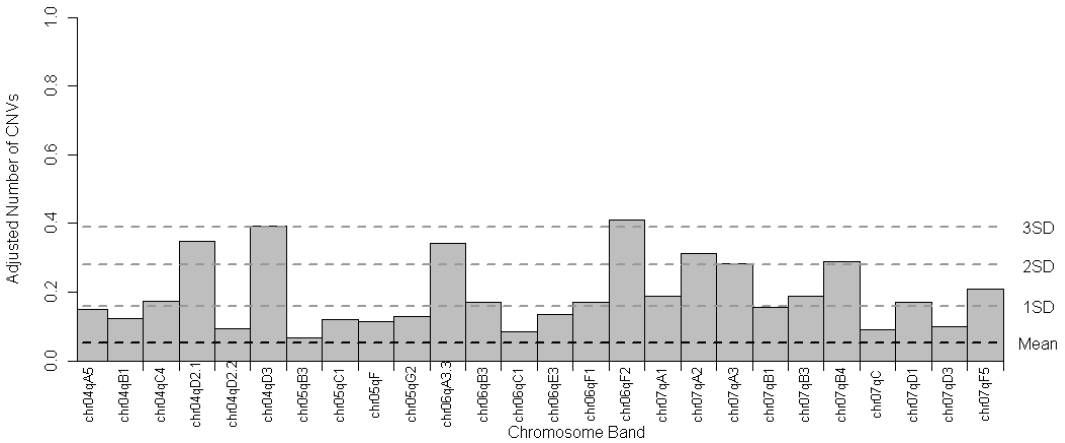
Figure 4.20 – De Novo CNVs by Chromosome Band

Plot of non-overlapping de novo CNVs per Mb plotted by chromosome band. Grey dotted line indicates mean, black dotted line indicates standard deviations above the mean

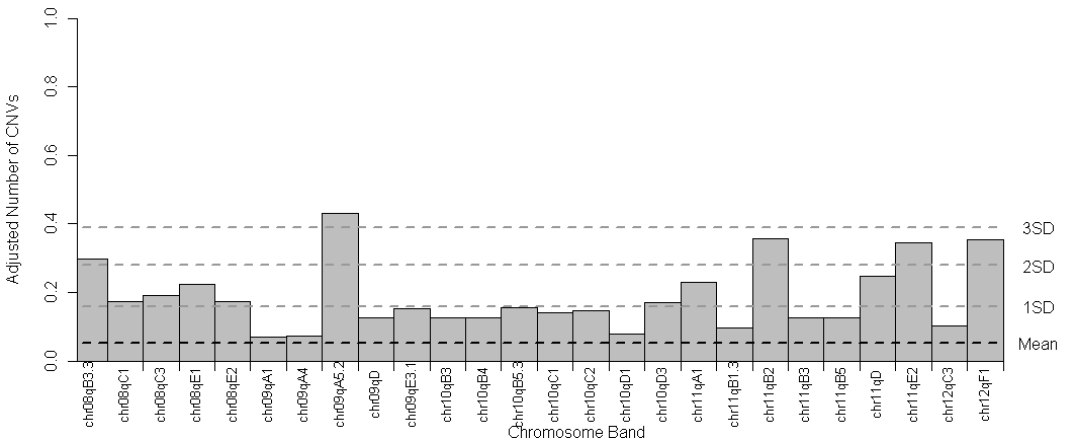
Chromosome 1-3



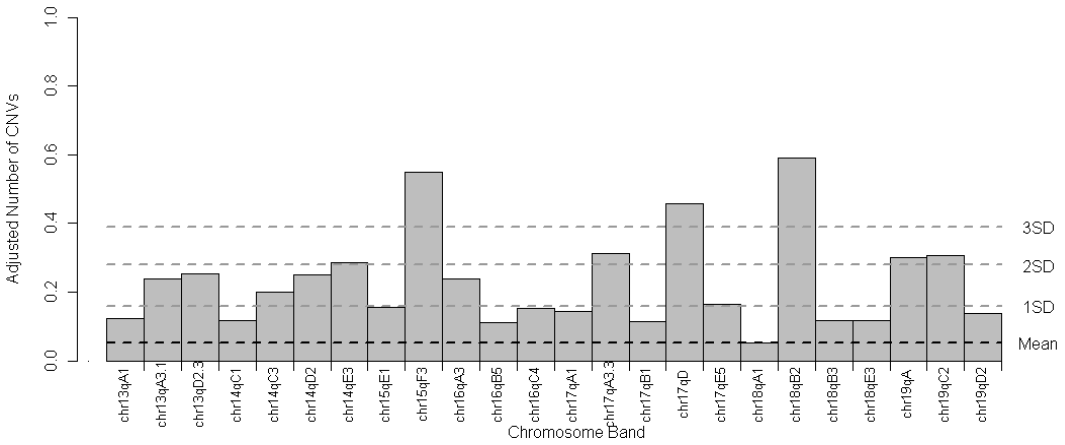
Chromosome 4-7



Chromosome 8-12



Chromosome 13-19



Comparing the CNVs identified in our sample with CNVs from other studies carried out on the C57BL/6 strain, we see some notable overlap in CNV locations (*Figure 4.21*, *Figure 4.22* and *Figure 4.23*), suggesting again that certain regions of the genome are particularly prone to duplications/deletions. Three CNVs overlapped with a previous paternal age study in mouse, one in chromosome 4 and two in chromosome 14 (Flatscher-Bader, Foldi et al. 2011). In Flatscher-Bader *et al's* study, these CNVs were only seen in the offspring of old fathers but we observe a less clear-cut distribution, with instances of these CNVs also seen in the offspring of young fathers. While the chromosome 4 CNV was only seen in the offspring of old fathers and not young fathers, we also observe it in two of the breeders suggesting this is not a *de novo* event linked to advanced paternal age. Furthermore, we see that both the CNVs on chromosome 14 occur in the offspring of both old and young fathers as well as some of the breeders. Both CNVs identified in the genome-wide study using low-resolution SNP arrays by Egan *et al* were also observed in our sample (Egan, Sridhar et al. 2007), as were four of the 38 CNVs identified between 14 colonies of C57BL/6 mice by Watkins-Chow *et al* (Watkins-Chow and Pavan 2008). These observations suggest that mice are likely to be polymorphic for CNVs at these locations (at least in the strains tested), and act to validate the array-CGH data produced in this study, indicating that we are detecting real changes in copy number that are seen in other studies. Gene expression changes associated with these CNVs is analysed in section 6.5.6.

Figure 4.21 - CNVs which Overlap with Flatscher-Bader et al

'All CNVs' track displays CNVs observed in our study

Figure 4.22 - CNVs which Overlap with Watkins-Chow et al

'All CNVs' track is CNVs observed in our study

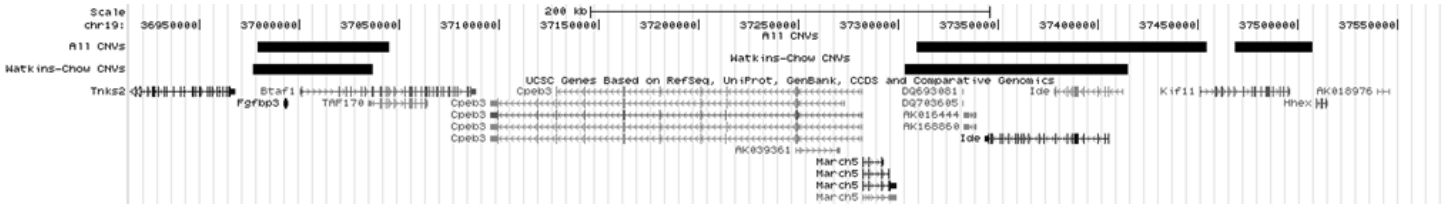
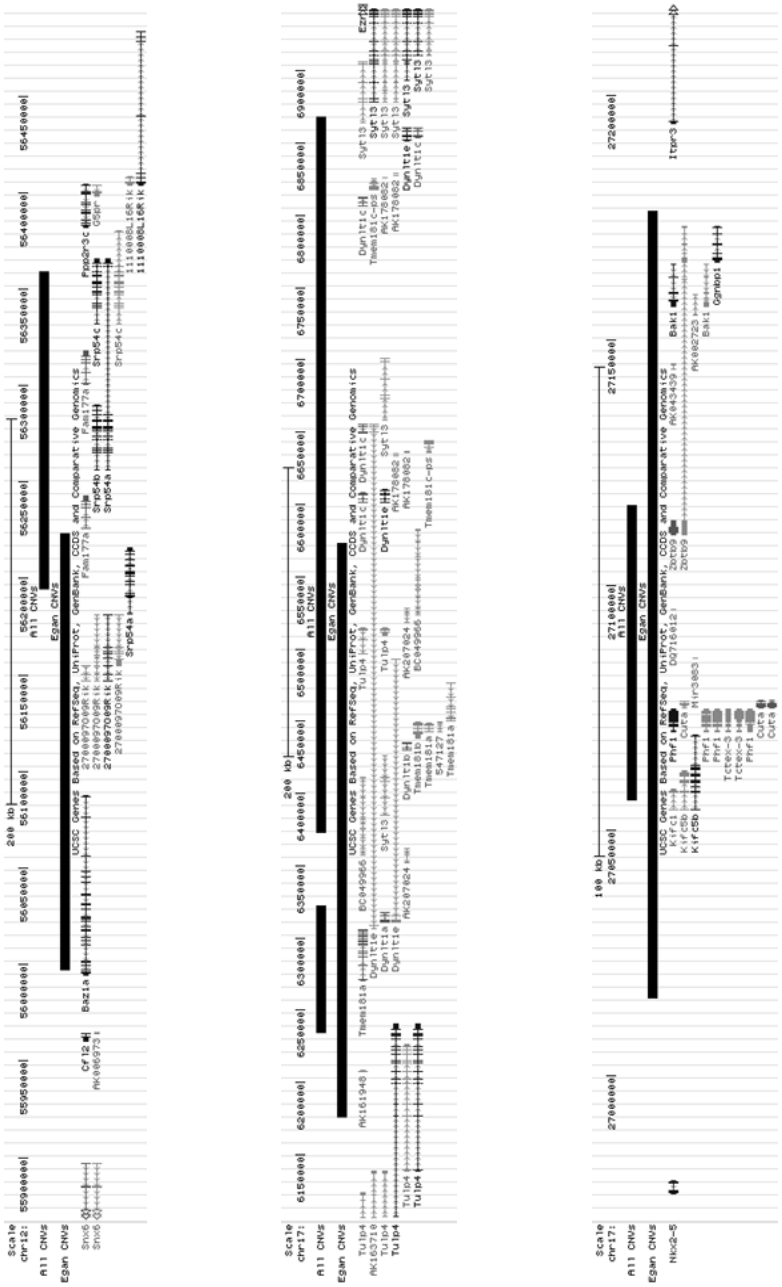


Figure 4.23 - CNVs which Overlap with Egan et al

'All CNVs' track is CNVs observed in our study

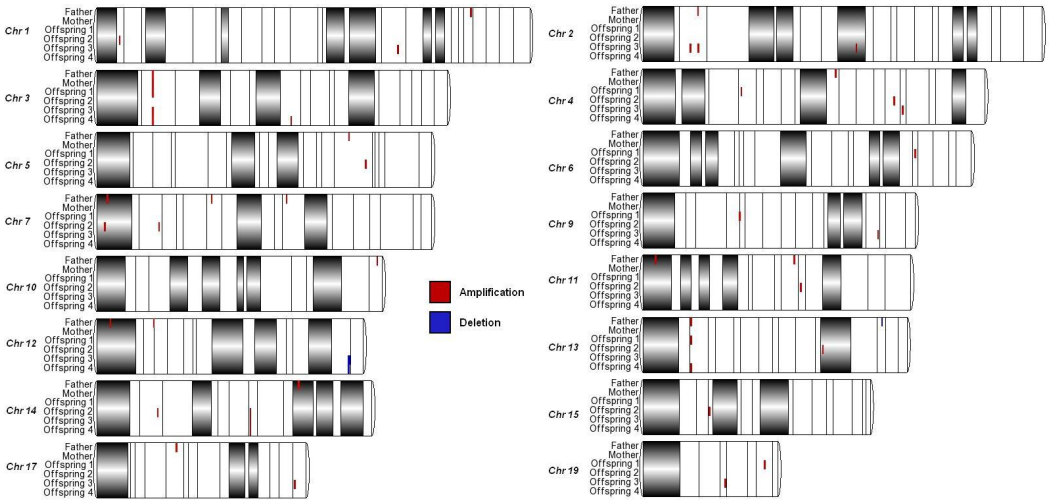


4.5.2 CNVs and Paternal Age

4.5.2.1 Total CNV Burden and Paternal Age

First we examined the affect of paternal age on total CNV burden. This analysis assessed inherited and *de novo* CNVs in all offspring (offspring of young fathers, n = 18; offspring of old fathers, n = 22) except PA02 which was excluded as an obviously extreme outlier. *Figure 4.24* shows a family of three male offspring plotted with their parents showing the inheritance of CNVs and the occurrence of all CNVs. CNVs in other families are shown in *Appendix 1*.

Figure 4.24 - Family 2 All CNV Karyotype from Partek



The average number of CNVs by group and family is shown in *Figure 4.25*. The total number of CNVs was not significantly different between the offspring of young and old fathers (t = -0.82, d.f. = 38, p = 0.42), although it was nominally higher in the offspring old fathers at 8.3 compared to 6.7 in the offspring of young fathers.

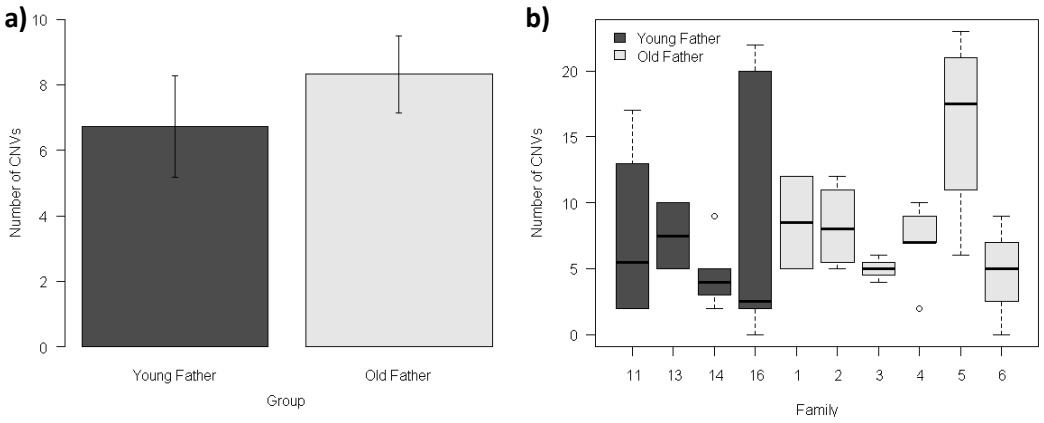


Figure 4.25 – Number of CNVs

a) By Group

b) By Family

The average size in kbps of the CNVs by groups and family is shown in *Figure 4.26*. There was no significant difference between groups in the average size of CNVs (t = 1.00, d.f. = 38, p = 0.33) with the average size in the offspring of young fathers being 99.6kbps and the average size in the offspring of old fathers being 84.5kbps.

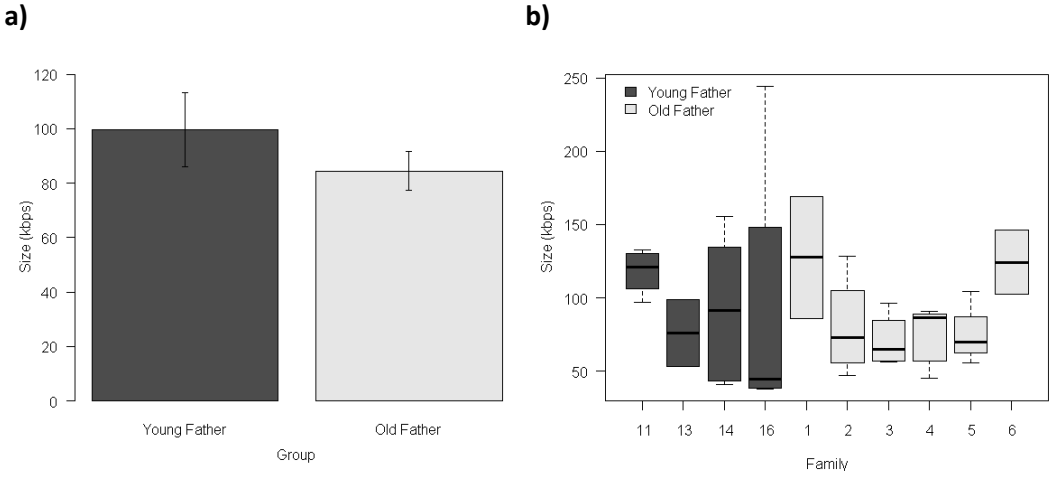


Figure 4.26 – Size of CNVs (kbps)

a) By Group

b) By Family

The average number of genes affected per CNV by each group and family is shown in *Figure 4.27*. There is no significant differences observed in the number of genes per CNV between the offspring of young and offspring of old fathers (t =1.48, d.f. = 38, p = 0.16) with the average in the offspring of young fathers being 3.0 genes per CNV and the average in the offspring of old fathers being 1.8 genes per CNV. The average in the offspring of young fathers is inflated by an outlier in family 16 (PA25) which had an average of 14 genes per CNV affected as one of the CNV is very large, spanning 94 genes.

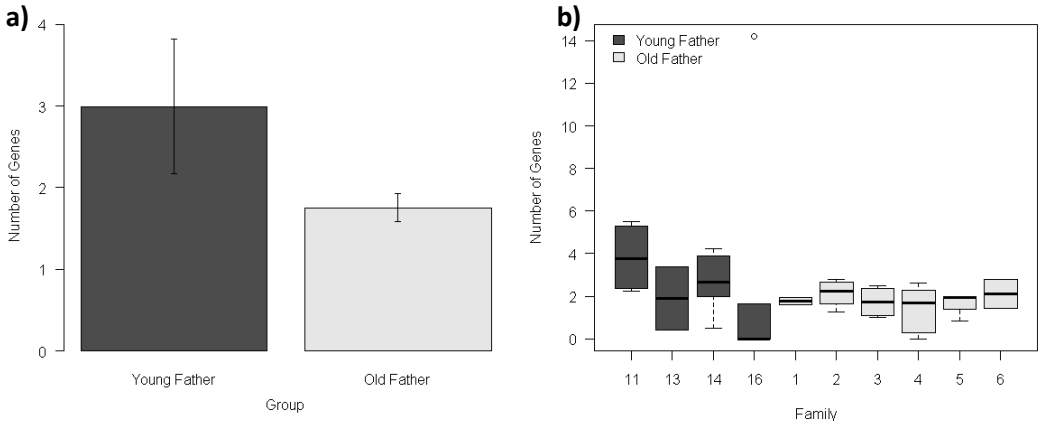


Figure 4.27 – Number of Genes per CNV

a) By Group

b) By Family

The total number of genes affected by the total burden of CNVs in an individual split by group and family is shown in *Figure 4.28*. There is no significant difference in the total number of genes between the offspring of young fathers and offspring of old fathers (t =1.00, d.f. = 38, p = 0.33) with the average in the offspring of young fathers being 31.8 and the average in the offspring of old fathers being 15.0. The large difference in total genes affected is again inflated by PA25 in family 16 who had 284 genes affected by CNVs.

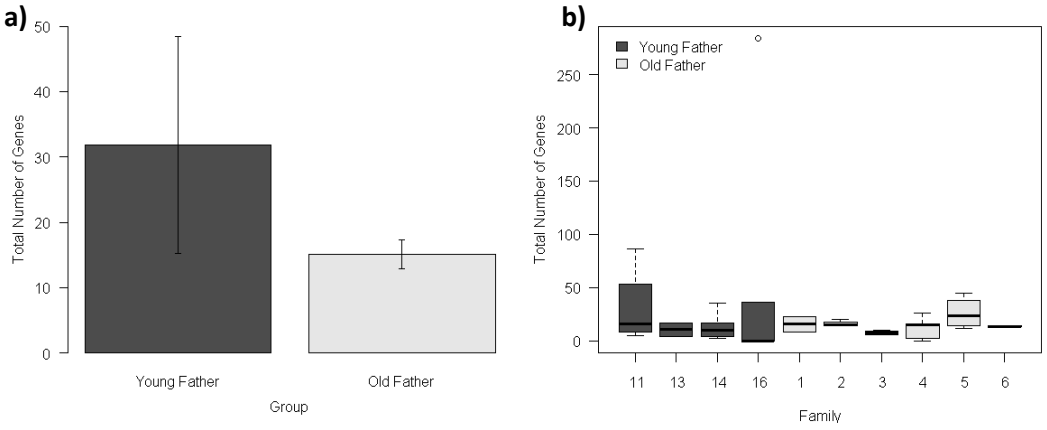


Figure 4.28 – Total Number of Genes affected by CNVs

a) By Group

b) By Family

The number of CNVs split by type (deletions or duplications) by groups and family is shown in *Figure 4.29*. There was no significant difference between the offspring of young fathers and of offspring of old fathers in the number of types of CNVs they have (duplications (t =-0.70, d.f. = 38, p = 0.49), deletions (t =-0.95, d.f. = 38, p = 0.35)). The average number of deletions and duplications in the offspring of young fathers was 0.5 and 6.7 respectively and the average number of deletions and duplications in the offspring of old fathers were 0.8 and 8.0 respectively.

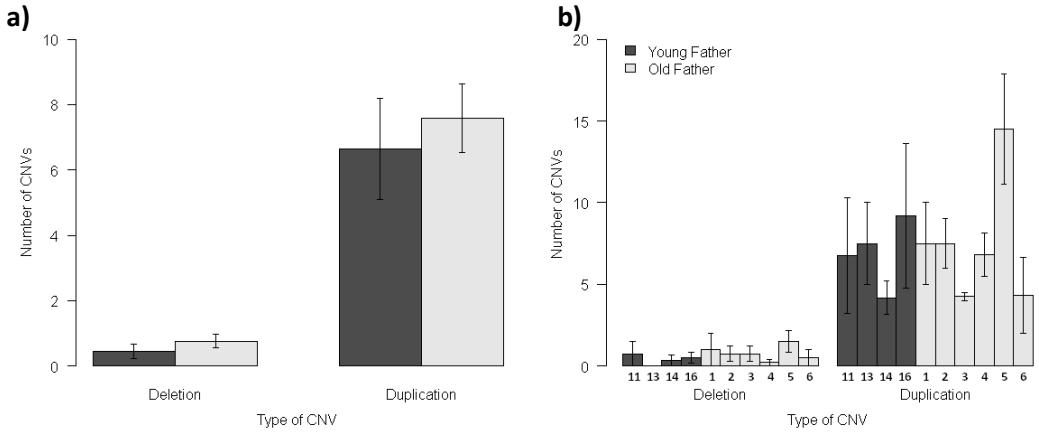


Figure 4.29 – Number of Deletions and Duplications

a) By Group

b) By Family

4.5.2.2 Inherited CNVs Ingenuity Pathway Analysis

The genes affected by all non-overlapping CNVs were subjected to IPA analysis to investigate networks which were enriched for CNVs with paternal age (*Table 4.4*). No networks associated with aging or the psychiatric conditions associated with paternal age were highlighted in the IPA analysis. Most of the genes highlighted in IPA were associated with gene expression. The genetic disorders highlighted are colon cancer and muscular dystrophy.

Table 4.4 – Networks identified in IPA of Gene Affected by CNVs

CNVs from all offspring

Genes	Top Functions
26	Genetic Disorder, Neurological Disease, Molecular Transport
25	Carbohydrate Metabolism, Molecular Transport, Small Molecule Biochemistry
25	Cell Morphology, Infectious Disease, Cell-To-Cell Signalling and Interaction
23	Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair
22	Cell Signalling, Nucleic Acid Metabolism, Small Molecule Biochemistry
21	Infectious Disease, Cell-To-Cell Signalling and Interaction, Haematological System Development and Function
20	Cellular Growth and Proliferation, Cell Death, Cellular Development
20	Antigen Presentation, Cell-To-Cell Signalling and Interaction, Haematological System Development and Function
16	Cell Morphology, Tissue Development, Cellular Movement
15	Cell-To-Cell Signalling and Interaction, Cellular Development, Nervous System Development and Function
15	Cell Signalling, Genetic Disorder, Neurological Disease
14	Embryonic Development, Organ Development, Organ Morphology

CNVs from offspring of young fathers

Genes	Top Functions
26	Gene Expression, Cellular Development, Protein Synthesis
24	Drug Metabolism, Endocrine System Development and Function, Lipid Metabolism
24	Cellular Assembly and Organization, DNA Replication, Recombination, and Repair, Post-Translational Modification
24	Free Radical Scavenging, Cell Death, Cellular Development
23	Cellular Development, Hematopoiesis, Cellular Growth and Proliferation
19	Cellular Assembly and Organization, Cellular Compromise, Cell-To-Cell Signalling and Interaction
16	Gene Expression, Lipid Metabolism, Nucleic Acid Metabolism
15	Free Radical Scavenging, Cellular Movement, Cellular Development
14	Cell-To-Cell Signalling and Interaction, Haematological System Development and Function, Hematopoiesis
14	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry
14	Genetic Disorder, Skeletal and Muscular Disorders, Developmental Disorder
13	Cell Cycle, Cellular Development, Connective Tissue Development and Function

CNVs from offspring of old father

Genes	Top Functions
20	Molecular Transport, Cancer, Gene Expression
18	Cellular Development, Hematopoiesis, Gene Expression
18	Endocrine System Disorders, Gastrointestinal Disease, Genetic Disorder
16	Carbohydrate Metabolism, Molecular Transport, Small Molecule Biochemistry
14	Cell-To-Cell Signalling and Interaction, Cellular Growth and Proliferation, Cell Death
13	Cell-To-Cell Signalling and Interaction, Nervous System Development and Function, Cell Morphology
7	Cell Signalling, Vitamin and Mineral Metabolism, Cell-To-Cell Signalling and Interaction
6	DNA Replication, Recombination, and Repair, Energy Production, Nucleic Acid Metabolism
1	Embryonic Development, Organismal Development, Cellular Development
1	Cell Cycle, Reproductive System Development and Function, Cell Death
1	Carbohydrate Metabolism, Cell-To-Cell Signalling and Interaction, Cellular Growth and Proliferation
1	Molecular Transport, RNA Trafficking, RNA Post-Transcriptional Modification

14	Post-Translational Modification, Genetic Disorder, Haematological Disease
14	Cellular Function and Maintenance, Connective Tissue Development and Function, Cellular Growth and Proliferation
14	Molecular Transport, Protein Synthesis, Protein Trafficking
13	Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization, Reproductive System Development and Function
13	Cell Signalling, Vitamin and Mineral Metabolism, Genetic Disorder
13	Cell Morphology, Cellular Function and Maintenance, DNA Replication, Recombination, and Repair
13	Cancer, Cell Cycle, Connective Tissue Development and Function
13	Cell Cycle, DNA Replication, Recombination, and Repair, Genetic Disorder
12	Endocrine System Development and Function, Lipid Metabolism, Molecular Transport
11	Carbohydrate Metabolism, Molecular Transport, Small Molecule Biochemistry
10	Behaviour, Gene Expression, Neurological Disease
10	Cellular Growth and Proliferation, Endocrine System Development and Function, Reproductive System Disease
8	Cell Morphology, Cardiovascular System Development and Function, Organismal Development

13	Cell Morphology, Cellular Assembly and Organization, Cellular Function and Maintenance
11	Carbohydrate Metabolism, Lipid Metabolism, Small Molecule Biochemistry
11	Gene Expression, Cell Cycle, Skeletal and Muscular System Development and Function
11	Genetic Disorder, Metabolic Disease, Neurological Disease
11	Cell-To-Cell Signalling and Interaction, Reproductive System Development and Function, Cellular Development
10	Cellular Development, Connective Tissue Development and Function, Haematological System Development and Function
10	Developmental Disorder, Genetic Disorder, Skeletal and Muscular Disorders
10	Carbohydrate Metabolism, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair
5	Post-Translational Modification, Protein Degradation, Protein Synthesis
1	Cell Cycle, Reproductive System Development and Function, Cell Death
1	Genetic Disorder, Immunological Disease, Cellular Development
1	Embryonic Development, Organismal Development, Organ Development
1	Post-Translational Modification, Cell-To-Cell Signalling and Interaction, Reproductive System Development and Function

1	Post-Translational Modification, Cell-To-Cell Signalling and Interaction, Reproductive System Development and Function
1	Antigen Presentation, Cell-To-Cell Signalling and Interaction, Haematological System Development and Function
1	Genetic Disorder, Metabolic Disease, Neurological Disease
1	Carbohydrate Metabolism, Drug Metabolism, Small Molecule Biochemistry
1	Cell Morphology, Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization

4.5.2.3 De Novo CNVS and Paternal Age

Data in this section includes array data from three families of the offspring of young fathers and three families from the offspring of old fathers for whom both parental and offspring high-quality CGH data was obtained. As discussed in section 4.5.1, data from families 16, 1, 4 and 6 were removed from *de novo* CNV analysis due to an unusual CNV profile in the breeders, potentially resulting from malignancy in the spleen. *De novo* CNVs from all individuals, split by group and family and arranged by chromosome, are shown in *Appendix 1*. As already described, *de novo* CNVs are generally distributed evenly throughout the genome with some exceptions (*Figure 4.20*). The number of *de novo* CNVs in both the offspring of old fathers and offspring of young fathers is shown in *Figure 4.30a,* and split into families in *Figure 4.30b*. There is no significant association with number of CNVs and paternal age (t = 0.06, d.f. = 18.68, p-value = 0.95) with the average number per individual in the offspring of young fathers being 5.5 and the average number per individual in the offspring of old fathers being 5.4.

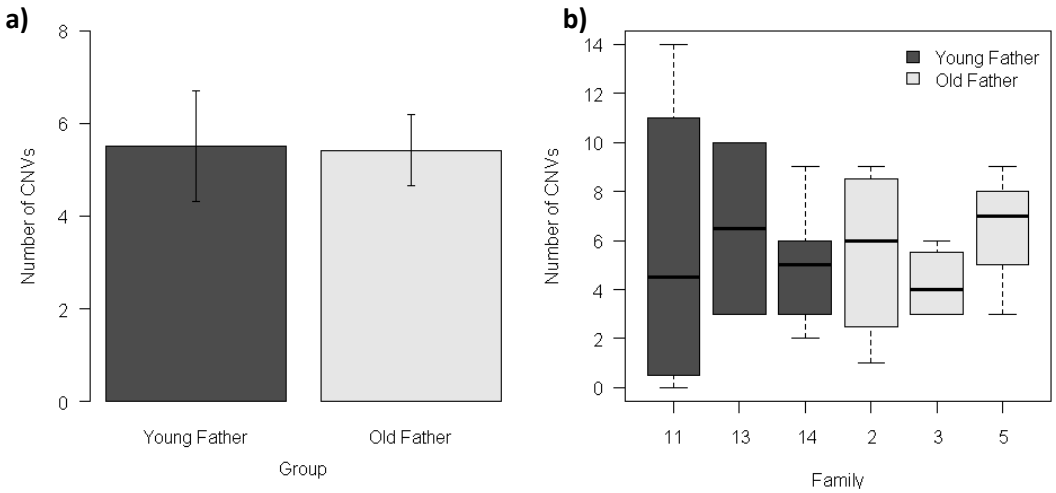


Figure 4.30 - Number of De Novo CNVs

c) By Group

d) By Family

The average size of *de novo* CNVs between the offspring of old fathers and offspring of young fathers is shown in *Figure 4.31a,* and presented split by family in *Figure 4.31b*. There is no significant association with number of CNVs and paternal age t = 0.60, d.f. =

14.65, p-value = 0.56) with the average in offspring of young fathers being 67.8kbps and in offspring of old fathers being 59.2kbps.

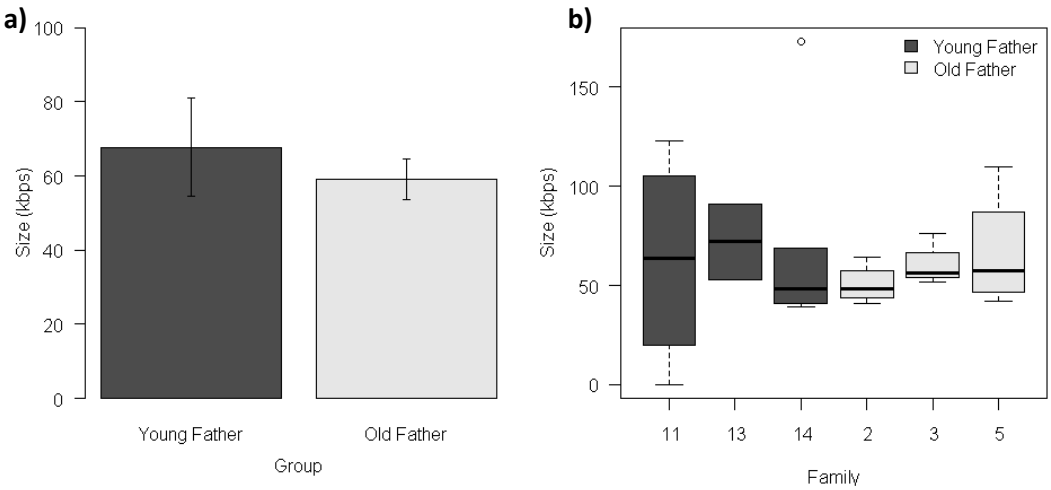


Figure 4.31 – Average Size of De Novo CNVs (kbps)

a) By Group

b) By Family

The average number of genes spanned by *de novo* CNVs in the offspring of old fathers and offspring of young fathers is shown in *Figure 4.32a,* and presented split by family in *Figure 4.32b*. There is no significant association between the number of CNVs and paternal age (t = 0.28, d.f. = 14.79, p-value = 0.78), with the average in offspring of young fathers being 2.1 and in offspring of old fathers being 1.9.

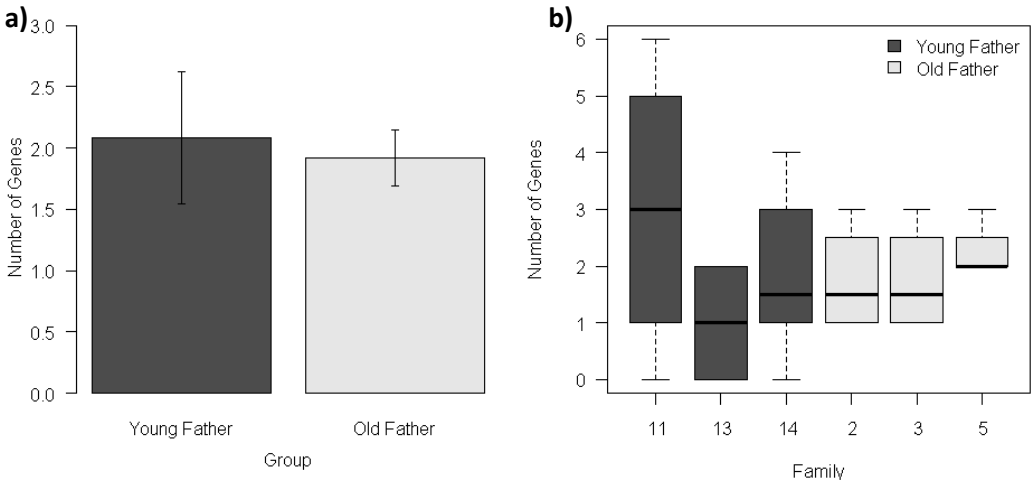


Figure 4.32 - Number of Genes per De Novo CNV

a) By Group

b) By Family

The number of genes affected by all CNVs per individual in the offspring of old fathers and offspring of young fathers is shown in *Figure 4.33a,* and presented split by family in *Figure 4.33b*. There is no significant association with number of CNVs and paternal age (t = 0.56, d.f. = 13.85, p-value = 0.58), with the average in offspring of young fathers being 12.8 and in offspring of old fathers being 9.8.

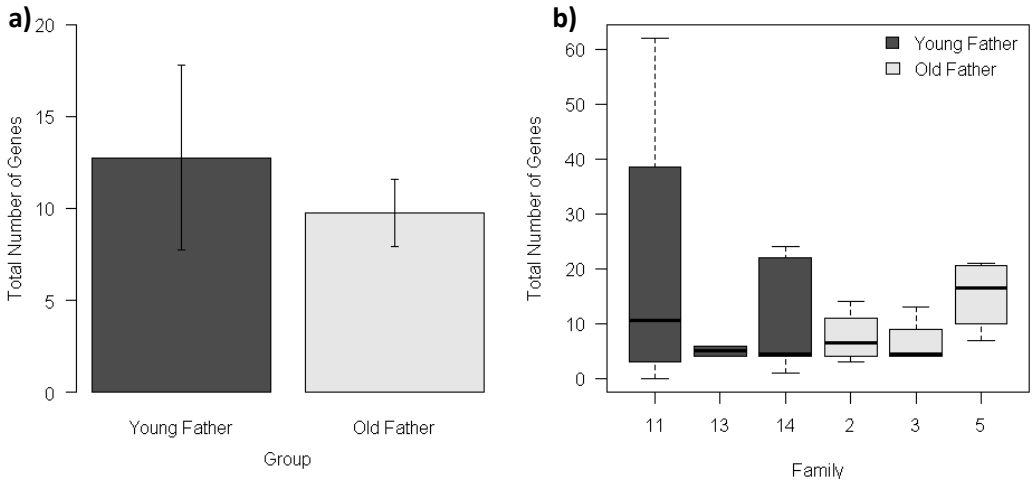


Figure 4.33 - Total Number of Genes Affected by CNVs across the Genome

a) By Group

b) By Family

The number of *de novo* deletion and duplication events between the offspring of old fathers and offspring of young fathers is shown in *Figure 4.34a,* and presented split by family in *Figure 4.34b*. No significant differences were observed between the groups in the number of duplications (t = -0.06, d.f. = 18.54, p-value = 0.95), with the average being 5.0 in the offspring of young fathers and 5.1 in the offspring of old fathers. Similarly, no significant difference in the number of deletions was observed (t = 0.48, d.f. = 18.91, p-value = 0.63), with the average in offspring of young fathers being 0.5 and in offspring of old fathers being 0.3.

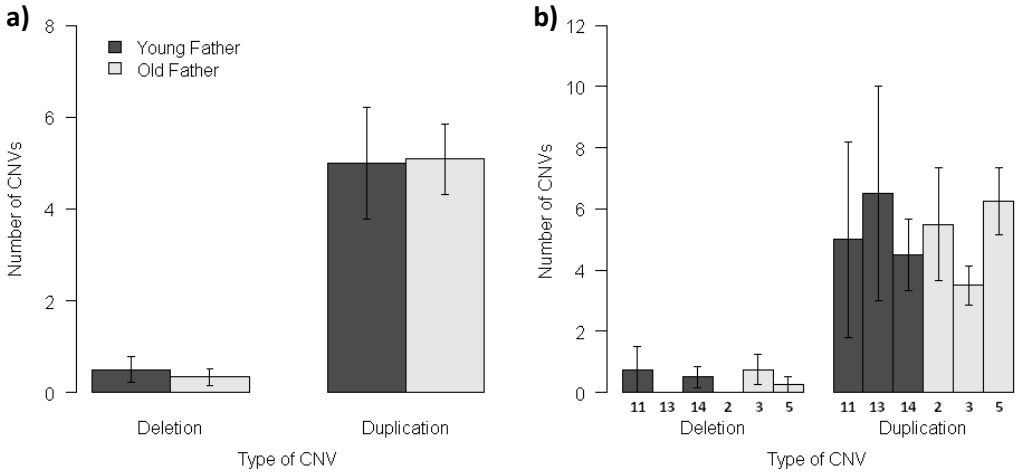


Figure 4.34 - Number of Deletion and Duplication De Novo CNVs

a) By Group

b) By Family

The number of *de novo* CNVs occurring by chromosome and the number of times they occur in each group is shown in *Table 4.5*. No association was seen with paternal age and number of *de novo* CNVs for copy number changes at any chromosome.

Table 4.5 – De Novo CNVs by Chromosome and Group

Chr	Chr Count	Young Chr Count	Old Chr Count	p-value	t-value
Chr1	10	5	5	1.00	0.00
Chr2	18	8	10	0.75	0.32
Chr3	5	3	2	0.63	-0.48
Chr4	9	4	5	0.69	0.41
Chr5	4	2	2	1.00	0.00
Chr6	6	3	3	1.00	0.00
Chr7	12	4	8	0.27	1.14
Chr8	4	1	3	0.30	1.08
Chr9	5	2	3	0.63	0.48
Chr10	7	4	3	0.72	-0.36
Chr11	7	4	3	0.75	-0.32
Chr12	5	3	2	0.70	-0.39
Chr13	7	3	4	0.75	0.32
Chr14	3	1	2	0.56	0.56
Chr15	3	1	2	0.56	0.56
Chr16	3	2	1	0.56	0.56
Chr17	10	6	4	0.50	-0.67
Chr18	5	2	3	0.63	0.48
Chr19	8	6	2	0.16	-1.48

The number of CNVs, genomic size, and number of genes affected were correlated with the behavioural measures discussed in (section 3.5), but no behavioural phenotypes were significantly correlated with any of these measures.

4.5.2.4 De Novo CNV Ingenuity Pathway Analysis

Genes affected by CNVs were subjected to Ingenuity Pathway Analysis (IPA) and the pathways identified for all *de novo* CNVs and in *de novo* CNVs by group using this approach is shown in *Table 4.6*. All networks were unrelated to aging or the psychiatric disorders of interest. Most genes affected by CNVs were in cell cycle and immune response.

Table 4.6 – Networks Identified in IPA of Gene Affected by De Novo CNVs

CNVs from all offspring

Genes	Top Functions
22	Cell Cycle, Cell Morphology, Carbohydrate Metabolism
22	Decreased Levels of Albumin, RNA Post-Transcriptional Modification, DNA Replication, Recombination, and Repair
16	Cellular Development, Hematopoiesis, Cell Signalling
15	Cellular Assembly and Organization, Cellular Function and Maintenance, Cellular Compromise
14	Metabolic Disease, Lipid Metabolism, Molecular Transport
14	Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization, Cellular Function and Maintenance
13	Cell Morphology, Connective Tissue Development and Function, Haematological System Development and Function
13	Lipid Metabolism, Nucleic Acid Metabolism, Small Molecule Biochemistry
12	Antigen Presentation, Cell-To-Cell Signalling and Interaction, Inflammatory Response
11	Cellular Development, Haematological System Development and Function, Hematopoiesis
9	Cell Death, Nervous System Development and Function, Genetic Disorder
8	Cellular Movement, Tumour Morphology, Cell Death
1	Cell Cycle, Reproductive System Development and

CNVs from offspring of young fathers

Genes	Top Functions
24	Cell Cycle, Cell-mediated Immune Response, Cellular Development
18	Cell Morphology, Cellular Development, Cellular Growth and Proliferation
16	Cellular Function and Maintenance, Lipid Metabolism, Molecular Transport
13	Cellular Growth and Proliferation, Cell Cycle, Cellular Development
12	Tissue Morphology, Cellular Development, Haematological System Development and Function
12	Molecular Transport, Protein Trafficking, Lipid Metabolism
1	Cellular Assembly and Organization, Post-Translational Modification, Molecular Transport
1	Cell Cycle, Reproductive System Development and Function, Cell Death
1	Cell Death, Cell Morphology, Digestive System Development and Function
1	Cellular Function and Maintenance, Dermatological Diseases and Conditions, Embryonic Development
1	Post-Translational Modification, Cell-To-Cell Signalling and Interaction, Reproductive System Development and Function
1	Cell Cycle, Reproductive System Development and Function, Cancer
1	Post-Translational Modification, Protein Synthesis,

CNVs from offspring of old fathers

Genes	Top Functions
21	Cell Death, Haematological System Development and Function, Gene Expression
16	Developmental Disorder, Neurological Disease, Cancer
12	Gene Expression, DNA Replication, Recombination, and Repair, Cell Morphology
11	Inflammatory Response, Cell Cycle, Cellular Function and Maintenance
9	Cellular Development, Nervous System Development and Function, Behaviour
1	Carbohydrate Metabolism, Cell-To-Cell Signalling and Interaction, Cellular Growth and Proliferation
1	Cell Death, Genetic Disorder, Skeletal and Muscular Disorders
1	Molecular Transport, RNA Trafficking, RNA Post-Transcriptional Modification
1	Antigen Presentation, Cell-To-Cell Signalling and Interaction, Haematological System Development and Function
1	Genetic Disorder, Metabolic Disease, Neurological Disease
1	Carbohydrate Metabolism, Drug Metabolism, Small Molecule Biochemistry
1	Cell Morphology, Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization
1	Infectious Disease, Cellular Assembly and

	Function, Cell Death
1	Cell Death, Genetic Disorder, Skeletal and Muscular Disorders
1	Molecular Transport, RNA Trafficking, RNA Post-Transcriptional Modification
1	Post-Translational Modification, Cell-To-Cell Signalling and Interaction, Reproductive System Development and Function
1	Antigen Presentation, Cell-To-Cell Signalling and Interaction, Haematological System Development and Function
1	Cell Cycle, Reproductive System Development and Function, Cancer
1	Post-Translational Modification, Protein Synthesis, Cellular Assembly and Organization
1	Genetic Disorder, Neurological Disease, Skeletal and Muscular Disorders
1	Cell Cycle, DNA Replication, Recombination, and Repair, Dermatological Diseases and Conditions
1	Cancer, Reproductive System Disease, Connective Tissue Disorders

	Cellular Assembly and Organization
1	Genetic Disorder, Neurological Disease, Skeletal and Muscular Disorders
1	Cell Cycle, DNA Replication, Recombination, and Repair, Dermatological Diseases and Conditions
1	Cellular Movement, Cellular Assembly and Organization, Cellular Function and Maintenance
1	Cancer, Reproductive System Disease, Connective Tissue Disorders
1	Cell Morphology, Cellular Assembly and Organization, Cellular Growth and Proliferation

	Organization, Cellular Compromise
1	Cellular Assembly and Organization, Lipid Metabolism, Small Molecule Biochemistry
1	Developmental Disorder, Genetic Disorder, Metabolic Disease
1	Carbohydrate Metabolism, Small Molecule Biochemistry, Cellular Assembly and Organization

4.6 Discussion

In this Chapter, I used CGH in combination with high-resolution microarrays to examine whether advanced paternal age is associated with an increased burden of CNVs in an inbred mouse model using the C57BL/6 strain. The analyses suggest that the offspring of older mice are not characterised by an increased incidence of *de novo* or total number of CNVs. Furthermore, the number of genes affected by CNVs and the proportion of deletions to duplications is not different between the offspring of young fathers and the offspring of old fathers. In this sample, therefore, there is no obvious association between advanced paternal age and the occurrence of *de novo* CNVs or their attributes. Furthermore, pathway analysis of the genes disrupted by the CNVs discovered highlights no obvious functional networks that can explain the paternal age behavioural differences seen in the old father group (see Chapter 3). Furthermore, by looking at the genes affected by CNVs in our sample, we see no overlap with the recurrent human CNVs implicated in ASD, schizophrenia or BD – three neuropsychiatric disorders epidemiologically linked to advanced paternal age and previously associated with an increased CNV burden. Since this study is about paternal age and the risk of psychiatric disease, we had hypothesised that some homologous regions of the genome may be affected by CNVs, especially given the social and exploratory behavioural deficits observed in these animals (see Chapter 3). As the mouse genome is not exactly homologous to the human genome in terms of sequence or the relative location of genes, and given that our behavioural measures did not map exactly on to human neuropsychiatric conditions, it may be that different loci are at risk for disruption in our model. Repetitive regions and regions close to centromeres and telomeres are expected to be more at risk of *de novo* CNVs in humans, however because mouse chromosome are acrocentric they have centromeres at the end of the chromosomes and so have different regions at risk to the mechanisms underlying structural variation. Unfortunately, because of their often highly repetitive nature, the regions most prone to CNVs are not often represented on the microarrays used in our experiments making comparisons with other studies difficult.

Of note was the overall number of CNVs observed in our sample regardless of paternal age given that we were using an inbred mouse strain (thus presumed to be 'genetically identical') obtained commercially from the same supplier. In addition to the actual burden of CNVs we observe considerable polymorphism across animals, which has obvious implications for rodent behavioural/environmental studies which are often assumed to be performed on

genetically-identical inbred animals. To date, only a few other studies of copy number variation in mice have been published (Egan, Sridhar et al. 2007; Watkins-Chow and Pavan 2008), but as described previously in this chapter, these also report evidence of polymorphic CNVs within inbred strains. Although the burden of CNVs (and rate of *de novo* events) we observe in our study is higher than in other mouse studies within C57BL/6, this may be a function of the more sensitive high-resolution array platform used in our study, and the limited number of animals assessed in any single study yet performed. In this study, we observe a higher incidence of duplications than deletions in both inherited (268:20) and *de novo* CNVs (121:10). This is contrary to previous studies in humans and mouse which report a more equal incidence of *de novo* duplications and deletions (Itsara, Wu et al. 2010) (Egan, Sridhar et al. 2007). Again, these differences could reflect differences in the platforms used, and the relative small sample sizes employed across all studies to date.

De novo CNVs have not previously been shown in human studies to be more frequent in the offspring of advanced age fathers, although chromosomal rearrangements and translocations have been shown to be more frequent in older male sperm (Thomas, Morris et al. 2010). It's therefore plausible that the increased burden of genomic abnormalities in the sperm of older males affects embryonic viability – of note, sperm cells from infertile males have been shown to contain higher rates of chromosomal aberrations compared to males in the general population (Chandley, Edmond et al. 1975; Testart, Gautier et al. 1996), and the same pattern is seen in males with a low sperm count (Retief, Van Zyl et al. 1984). Even if sperm containing chromosomal abnormalities do manage to fertilize an oocyte, the resulting zygote may not be viable due to the rearrangements or anomalies being lethal. Future studies will be carried out to analyse CNVs and genomic rearrangements in the sperm cells from the breeders themselves to assess whether these are characterised by an increased burden of structural variation.

Only one previous study has directly studied CNVs in relation to paternal age in a mouse model previously. In their study, Flatscher-Bader *et al* (2011) found seven *de novo* CNVs in a sample of six offspring sired by 2 advanced aged fathers (12 – 18 months old) compared to no *de novo* CNVs in the six offspring of 2 three month old fathers. However, two of these were located in the same gene adjacent to each other and so likely to be part of the same CNV.

Further analysis of these six CNVs did not replicate the observation with advanced paternal age in an independent same of mice (Flatscher-Bader, Foldi et al. 2011). This study does not, therefore, give convincing evidence for the role of *de novo* CNVs in relation to advanced paternal age – the results could be due to one family in the older father group having an abundance of CNVs compared to all families, not just compared to the offspring of younger fathers. Taken together with the results from our study, the evidence supporting an abundance of CNVs in the offspring of older male mice is currently limited.

The CGH microarray analyses presented in this chapter indicates that a relatively large number of CNVs were detected across all groups of animals, given that we utilised inbred mice derived from the same commercial breeding stock. A previous study using the Nimblegen array-based CNV detection method concluded that approximately one third of the CNVs detected by the 2.1M array were likely to be false positives, and it was also speculated that the false negative rate could be more than 50% (Agam, Yalcin et al. 2010).The problem of false positives and negatives is a potential confounder in many other studies of CNVs in human and mouse. Analysis of the same samples has yielded different findings with regard to both the number and location of CNVs in human studies (Conrad, Andrews et al. 2006; Locke, Sharp et al. 2006; McCarroll, Hadnott et al. 2006; Redon, Ishikawa et al. 2006), and in inbred mouse studies there is little overlap in the CNVs reported in the strains (Cutler, Marshall et al. 2007; Henrichsen, Vinckenbosch et al. 2009). Such inconsistencies could result from differences in the platforms used to discover CNVs or the types of analysis and normalization used in the data analysis stages. No standard definitions or protocols are available for the detection or analysis of CNVs, and so results are left to the interpretation of individual research groups (Scherer, Lee et al. 2007). For example, it has been suggested that the 23,897 CNVs identified in a recent analysis of healthy human control samples is likely to be an over-estimate due to the arbitrary approaches used to define CNV boundaries, experimental artefacts and incomplete quality control without verification (Alaerts and Del-Favero 2009). This variation makes CNV detection and investigation more complicated than other genomic approaches (e.g. SNP genotyping, gene expression or DNA methylation analysis, the other major methodologies utilised in this thesis). Due to this, the amount of false CNVs identified in this study is unknown, and future work will focus on verifying specific examples. Such technical errors, however, are likely to be largely consistent across samples and should not alter the overall conclusions of the analyses. Furthermore, it is reassuring that many of the CNVs shown

to recurrently occur in the mouse genome in previous studies are observed in our data (see *Figure 4.21*, *Figure 4.22* and *Figure 4.23*), indicating that we were detecting real changes in copy number.

A potential reason for the higher number of CNVs we observe compared to other studies could be the tissue we chose to assess. We did not measure the physiological properties of the spleen when dissecting the animals and so it is possible the spleen in test animals may have had a spleen disorder resulting in abnormal cells with a large number of chromosomal rearrangements. This is perhaps reflected in the unusual CNV profiles generated for the samples removed from the analyses presented here. These samples passed all QC metrics, but were characterised by an unusually large burden of structural variation. As discussed above, there is evidence that malignant disorders such as lymphoma are potentially common in the spleens of certain strains of inbred mouse (Murphy 1968; Brayton, Treuting et al. 2012), and a recent review of diseases in inbred mouse strains concluded that the most common causes and contributors to death in studies of aging C57BL/6J mice are lymphoma and hematopoietic neoplasms, suggesting that genomic alterations in the spleen and related tissues may occur widely with aging in this strain (Brayton, Treuting et al. 2012). Future studies should assess changes occurring in other tissues (for example the brain), in addition to germline (sperm) cells from the fathers. As discussed previously, CNVs can be subject to tissue specificity via mosacism (Locke, Sharp et al. 2006; Bruder, Piotrowski et al. 2008) and so our use of DNA isolated from spleen could limit the interpretation of our data, especially given the predominance of neuropsychiatric phenotypes in the context of advanced paternal age. Finally, because we did not use sperm cells from the fathers themselves, we are unable to estimate the actual rate of CNV mutations in the germline of older males.

We were somewhat restricted with regard to the number of animals we were able to assess in this analysis. Although we had access to the offspring of three different age fathers, only two ages were used in the CNV analysis presented here. As the third group were from fathers of 12 months old, it is plausible that the burden of CNVs may have been larger in this group. No DNA was available from the dam from family 16, and so the number of *de novo* CNVs from this family could not be completely ascertained as some CNVs from the offspring may have been inherited from the mother.

4.6.1 Conclusions

The results of this study and the lack of evidence to the contrary from other studies, indicates that there is no convincing evidence for an association between the number of *de novo* CNVs or the occurrence of specific *de novo* CNVs and advanced paternal age in mouse. Although previous research has demonstrated that the sperm from older males shows higher rates of chromosomal abnormalities than that from younger males, our data (and analogous epidemiological studies in humans) suggest that these do not seem to transfer to the offspring, possibly because the mutations are lethal or confer infertility. The results of this chapter indicate that it is unlikely that the paternal age effect is mediated by CNVs, particularly the behavioural effect observed in the same animals in Chapter 3. Expression changes associated with CNVs in this chapter will be discussed in section 6.5.6.

Chapter 5 - Epigenetic Differences Associated with Advanced Paternal Age in a Mouse Model

5.1 Abstract

Epigenetic changes have been observed in human studies of autism, schizophrenia and BD and have also been shown to accumulate over the life of an individual. One potential mechanism for paternal age effect is the transmission of age-associated epigenetic changes in the sperm of older fathers to their offspring. Epigenetic reprogramming resets the methylation status of the genome during gametogenesis and embryogenesis, but DMRs associated with imprinted loci and transposable elements appear to partially escape this genome wide resetting, indicating that they could be a potential vehicle for the transmission of methylation marks across generations. To investigate this hypothesis, global DNA methylation levels and DNA methylation across DMRs associated with 19 brain-expressed imprinted loci were assessed in the offspring of young, old and very old fathers in multiple tissues. The offspring of older fathers showed higher global methylation in DNA from the spleen and cerebellum. Three imprinted loci (Nesp, Kcnq1ot1 and Mcts2) showed consistently significant DNA methylation differences between the offspring of old fathers and young fathers in the cerebellum, and more subtle differences were observed at other DMRs. These results suggest it is plausible that the mechanism for the paternal age effect could be the transmission of age-accrued methylation marks from sperm of older fathers to their offspring.

5.2 Introduction

5.2.1 The Dynamic Epigenome

As described in more detail in Chapter 1 (section 1.6), epigenetic mechanisms control the reversible regulation of gene expression mediated principally through changes in DNA methylation and chromatin structure occurring independently of the DNA sequence (Henikoff and Matzke 1997). Due to its labile nature (Rakyan, Blewitt et al. 2002), the epigenome represents a good candidate for mediating interactions between genetic variation and the environment. Its dynamic, changeable nature has been demonstrated in numerous studies looking at DNA methylation changes associated with a range of environmental exposures (section 5.2.2), and ultimately provides a potential target for the development of novel therapeutic treatments. Environmental influences on the epigenome have been shown to be particularly important during gestation but can also occur periconceptually and later in life (Sinclair, Allegrucci et al. 2007). Once established, epigenetic marks can be maintained stably

through mitosis down cell lineages, potentially mediating the later-life manifestation of phenotypic effects (and diseases) caused by exposure to environmental factors early in development.

5.2.2 Epigenetics and Environmental Influences

A diverse range of environmental factors including diet, stress, air pollutants and non mutagenic toxic agents have been shown to influence the epigenome, both globally and at specific gene loci (reviewed in (Arita and Costa 2011)). For example, prenatal exposure to dietary factors can influence epigenetic processes, often leading to extreme changes in phenotype. DNA methylation at an IAP element upstream of the Agouti gene, for example, can be directly influenced by the intake of food supplemented with methyl donors (Dolinoy, Huang et al. 2007). Pregnant dams fed with a diet rich in methyl donors produce offspring with a high level of DNA methylation across this region, silencing expression from the cryptic promoter generated by the IAP element and resulting in a wild-type phenotype (brown fur). In contrast, pregnant dams fed with a diet low in methyl donors produce offspring with low levels of DNA methylation at this locus, resulting in offspring with a yellow coat (*Figure 5.1*), obesity and physiological deficits (Waterland and Jirtle 2003). Another example of the influence of prenatal diet on DNA methylation is observed in individuals exposed *in utero* to the Dutch Winter famine. This famine has been associated with hypomethylation across the *IGF2* imprinting DMR relative to same-sex siblings not exposed to prenatal famine, detected 60 years after the famine occurred (Heijmans, Tobi et al. 2008).

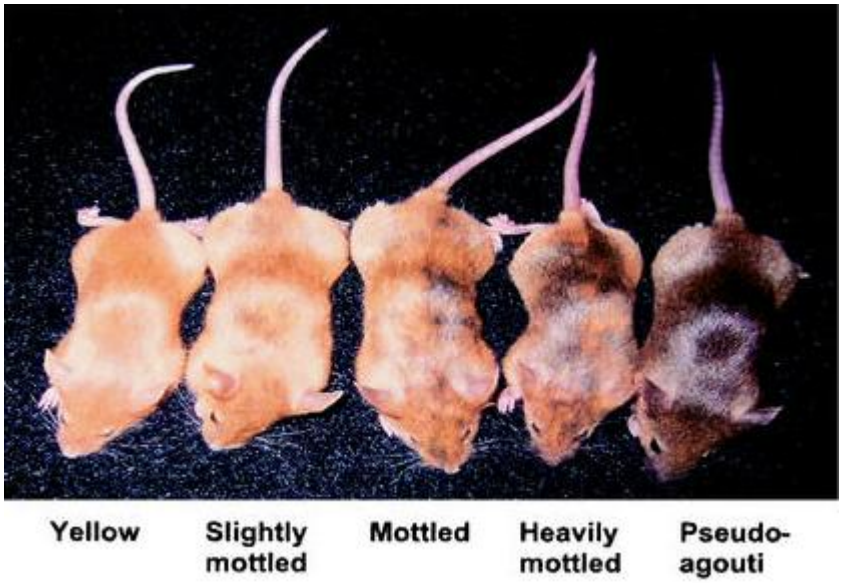


Figure 5.1 – Coat Colour Phenotype of Avy/a Offspring

Animals representing the five coat colour classes used to classify phenotype in the agouti viable yellow mouse model. The Avy alleles of yellow mice are hypomethylated, allowing for high expression of the mutant agouti gene. Avy hypermethylation in pseudo-agouti animals silences agouti expression.

Maternal care has also been demonstrated to affect DNA methylation in the offspring. In rats, for example, the offspring of high licking and grooming mothers have a significantly lower level of hippocampal DNA methylation at exon 1 of the glucocorticoid receptor gene than the offspring of low licking and grooming mothers (Weaver, Cervoni et al. 2004). This hypomethylation in the offspring of high licking and grooming mothers can be reversed in the adult offspring by the infusion of methyl donors (Weaver, Champagne et al. 2005). Early life stress, simulated in mice as separation of pups from the mother, also results in pups being hypomethylated at a CpG island upstream of the arginine vasopressin gene, which persists into adulthood (Murgatroyd, Patchev et al. 2009). From these examples, it can be seen that a diverse range of environmental influences can have a noticeable effect on DNA methylation that can stably alter gene expression and phenotype over development.

To date, it is not clear whether environmental influences on the father occurring preconception can affect his offspring via the meiotic transmission of epigenetic changes

occurring in the germline, although preliminary evidence suggests that such effects may occur. Mice exposed to particulate air pollution were found to be hypermethylated in their sperm compared to non-exposed mice (Yauk, Polyzos et al. 2008). Furthermore, gene-specific DNA methylation was altered in sperm cells in mice exposed to vinclozolin (a common dicarboximide fungicide), and strikingly this effect appears to be transmitted to subsequent male offspring (Anway, Cupp et al. 2005) over several generations suggesting that epigenetic changes in the male germline may be potentially inherited transgenerationally. More information about the evidence supporting transgenerational epigenetic inheritance is given in section 1.6.3. Although current evidence is limited, it is plausible that the sperm of older males may show age-related epigenomic changes caused by the accumulation of developmentally-, stochastically- or environmentally-induced epimutations that are subsequently transmitted to the offspring. This in turn could lead to the developmental manifestation of phenotypic differences, potentially related to the onset of disease. The research presented in this chapter examines whether epigenetic differences can be observed between the offspring of old and young fathers as a first step in exploring this hypothesis.

5.2.3 Epigenetics and Psychiatric Disease

Aetiological studies of ASD, schizophrenia and bipolar disorder have focused primarily on the interplay between genetic and environmental risk factors. Although twin and adoption studies highlight a clear inherited component to these disorders, genetic studies are often characterized by non-replication, small effect sizes, and significant heterogeneity. Several epidemiological, clinical, and molecular peculiarities associated with these psychiatric conditions are difficult to explain with traditional gene- and environment-based approaches, including the non-complete concordance between monozygotic twins, sexual dimorphism, peaks of susceptibility to disease coinciding with major hormonal rearrangements, parent-of-origin effects, and of particular relevance to this thesis, paternal age effects (Smith and Mill 2011). These observations have led to widespread speculation about the role of non-genetic factors, and influence of epigenetic variation in the development of these diseases has been a recent focus for research. Epigenetic changes associated with the three psychiatric disorders strongly linked to advanced paternal age (autism, schizophrenia, and bipolar disorder) are briefly reviewed in this section.

Autism: Although studies of epigenetic changes in autism are in their infancy, DNA methylation differences at several specific loci have been shown in autism. For example, increased DNA methylation at the *MECP2* promoter has been observed in autism cases compared to matched controls leading to decreased *MECP2* expression (Nagarajan, Hogart et al. 2006). This lower transcription of *MECP2* leads to decreased expression of *UBE1A* and *GABRB3* (Samaco, Hogart et al. 2005). Increased DNA methylation has also been shown in the promoter of the oxytocin receptor gene in autistic individuals compared to control samples (Gregory, Connelly et al. 2009). In a recent study of MZ twins discordant for autism, 73 separate loci showed differential methylation between discordant twins (Nguyen, Rauch et al. 2010).

Schizophrenia: Several candidate gene studies have uncovered epigenetic changes associated with schizophrenia. For example, the CpG island of *SOX10* has been shown to be hypermethylated in post-mortem brain samples from schizophrenia patients compared to controls (Iwamoto, Bundo et al. 2005). *RELN* has also been shown in a number of studies to be hypermethylated across its promoter in schizophrenia patients (Abdolmaleky, Cheng et al. 2005; Grayson, Jia et al. 2005; Torrey, Barci et al. 2005), and *COMT* exhibits hypomethylation at its promoter in schizophrenia patents compared to controls (Abdolmaleky, Cheng et al. 2006). Another candidate gene study focussing on the promoter of *5HTR1A* found increased methylation in schizophrenia patients than in control samples (Carrard, Salzmann et al. 2011). In a genome-wide case-control study using post-mortem frontal cortex tissue, 83 separate loci showed significant DNA methylation differences between schizophrenia cases and controls, including several loci related to brain development and neurobiological function (Mill, Tang et al. 2008). In a study of MZ twins discordant for schizophrenia, DNA from patients had a methylation profile across the *DRD2* promoter which was more similar to the affected twin from a different twin pair than their own unaffected co-twin (Petronis, Gottesman et al. 2003). Finally, a recent genome-wide methylation study looking at MZ twins discordant for schizophrenia reported the largest disease-associated differences between affected and unaffected twins to be at CpG sites in *PUS3, SYNGR2, KDELR1, PDK3, PPARGC1A, ACADL, FLJ90650* and *TUBB6* (Dempster, Pidsley et al. 2011).

Bipolar disorder: BD has been epigenetically assessed both individually and combined with schizophrenia in the context of "major psychosis". Again, many studies have focussed on specific candidate genes. *COMT* was found to exhibit promoter hypomethylation in cases compared to controls (Abdolmaleky, Cheng et al. 2006) and *5HTR1A* shows increased methylation at its promoter in BD cases compared to controls (Carrard, Salzmann et al. 2011). In a genome wide study 36 loci showed consistently significant DNA methylation differences between brain samples from affected and unaffected samples (Mill, Tang et al. 2008). A genome wide study in discordant MZ twin-pairs showed that the top eight loci exhibiting differences between affected and unaffected twin pairs were *GPR24, TLE6, STAB1, PPYR1, CTNNA2, ST6GALNAC1, C1orf35* and *IQCH* (Dempster, Pidsley et al. 2011). Looking at candidate genes in discordant MZ twins, *PIP5KL1, ARMC3, SMS* and *PPIEL* were significantly different between the affected twin and unaffected co-twin. *SMS* and *PPIEL* remained significantly different when looking at the data in a case-control analysis (Kuratomi, Iwamoto et al. 2008). Finally, looking at female MZ twin-pairs both discordant and concordant for bipolar disorder, discordant female twin pairs showed a greater degree of skewed X-chromosome inactivation than concordant twin pairs (Rosa, Picchioni et al. 2008).

5.2.4 Epigenetics and Aging

Epigenetic changes have been demonstrated to accumulate over the life of an individual, and recent studies suggest that aging is associated with profound effects on DNA methylation. Studies on MZ twins have demonstrated that MZ twins have a higher within-pair epigenomic correlation than DZ twins (Boks, Derks et al. 2009; Kaminsky, Tang et al. 2009; Wong, Caspi et al. 2010), but with age the epigenetic similarity between MZ twins appears to decrease (Fraga, Ballestar et al. 2005). DNA methylation at multiple genes has been shown to be strongly correlated with age (Boks, Derks et al. 2009). In Christensen *et al*'s genome wide methylation analysis across multiple tissues, over 300 CpG loci showed age-related DNA methylation alterations including 38 age related methylation differences in the brain (Christensen, Houseman et al. 2009). Bocklandt *et al* identified 88 sites that are correlated with age in DNA from saliva, 73 of which were located in CpG islands. Pathway analysis on the 80 genes associated with the 88 sites showed enrichment for genes involved in cardiovascular disease, neurological disease, and genetic disease (Bocklandt, Lin et al. 2011). In addition, Hernandez *et al* identified ten loci with significant genome-wide association with age across multiple brain regions (frontal cortex, temporal cortex, pons and cerebellum) (Hernandez, Nalls et al.

2011). Of relevance to studies of paternal age, male germline cells (sperm) also show age related epigenetic changes, with DNA methylation in a number of genes correlating with the age of the donor (Flanagan, Popendikyte et al. 2006). This age-related epigenetic drift in sperm leads to increased inter- and intra-individual variation in DNA methylation at multiple gene promoters (Flanagan, Popendikyte et al. 2006). Taken together, these studies suggest there is considerable epigenetic drift with age in many different tissues, including in the brain and germline.

In terms of my research on paternal age, an interesting hypothesis is that in the sperm of older males the maintenance of genomic imprints breaks down over the course of multiple cell divisions in spermatogenesis (Malaspina, Reichenberg et al. 2005). In other constantly dividing cell types, subsequent cell divisions do appear to alter the pattern of imprinting; for example, DNA methylation at imprinted loci increases with age in colon cells (Issa, Vertino et al. 1996). Given the evidence for incomplete epigenetic reprogramming in gametogenesis/embryogenesis (discussed in section 1.6.3), it is plausible that some of these age-associated epigenetic changes get transmitted to the offspring of older fathers. In this study we assessed global levels of DNA methylation and DNA methylation across DMRs associated with multiple brain-expressed imprinted genes in an animal model of paternal age.

5.2.5 Potential Mechanisms for Epigenetic Inheritance

Evidence for epigenetic inheritance down the male germline has been observed in rats exposed to fungicides and pesticides, where epigenetic changes occur in F1 populations and subsequent male offspring (Anway, Cupp et al. 2005). For the paternal age effect to be mediated by the transmission of age-acquired epigenetic variation from fathers to the offspring, we have to focus on genomic elements that potentially escape epigenetic reprogramming during germ cell development. There are several potential targets for the epigenetic changes (be they environmentally-, stochastically-, or developmentally-induced) associated with paternal age. These include the *cis*-acting regulatory elements of imprinted genes and CpG rich regions spanning house-keeping gene promoters and transposable elements such as LINE-1 and Alu elements (Dolinoy and Jirtle 2008). In this study, my focus is on epigenetic changes occurring at imprinted genes and transposable elements.

5.2.5.1 Imprinted Genes

As one allele is already inactive (section 1.6.2), imprinted genes are at a greater risk than non-imprinted genes of inactivation by mutations, loss of heterozygosity and epigenetic alterations to gene expression, leading to potentially pathogenic effects (Dolinoy and Jirtle 2008). For example, hypomethylation of the *IGF2* DMR leads to biallelic expression and is a potential risk factor in the development of colorectal cancer (Cui, Cruz-Correa et al. 2003). Since sperm undergo an increasing number of cellular divisions with age, it is plausible that there may be a break-down in the maintenance of genomic imprinting signatures that are subsequently transmitted to the offspring. Evidence for such a mechanism has emerged from studies investigating children conceived by *in-vitro* fertilization (IVF) and other forms of assisted reproductive technology (ART). For example, some cases of children conceived by intracytoplasmic sperm injection developing Angelman syndrome show a sporadic imprinting defect at the *SNRPN* loci (Cox, Burger et al. 2002; Orstavik, Eiklid et al. 2003). Offspring suffering from Beckwith-Wiedemann syndrome conceived with ART have been shown to have loss of imprinting of *KCNQ1OT1* (DeBaun, Niemitz et al. 2003; Gicquel, Gaston et al. 2003; Maher, Brueton et al. 2003) and *H19* (DeBaun, Niemitz et al. 2003). Although the reasons for the use of ART in these cases are unknown, these observations highlight the susceptibility of imprinted loci to epigenetic alterations, resulting in conditions with neuropsychiatric phenotypes.

As discussed in section 1.6.3, some regions of the genome escape complete epigenetic reprogramming during gametogenesis/embryogenesis (Lane, Dean et al. 2003; Popp, Dean et al. 2010), including some DMRs associated with imprinted genes and highly-repetitive sequences (*Figure 1.11*). As imprinted genes may also be particularly susceptible to epigenetic changes over the life-course in response to stochastic, developmental and environmental factors (Jirtle and Skinner 2007; Heijmans, Tobi et al. 2008), potentially leading to an accumulation of epigenetic variability with age in sperm (Flanagan, Popendikyte et al. 2006), the observation that some imprinting DMRs escape complete epigenetic reprogramming (Popp, Dean et al. 2010) suggests they could mediate the transmission of some paternal-age-induced epigenetic changes to subsequent generations. Of note, a recent study showed that DNA methylation in newborn babies is correlated with both maternal and paternal age at numerous gene promoters, including 12 known imprinted loci (Adkins, Thomas et al. 2011). Furthermore, in a study assessing imprinting errors in offspring, seven out of 17 cases showed

an identical alteration in DNA methylation detectable in the father's sperm (Kobayashi, Hiura et al. 2009), providing evidence that epigenetic inheritance may be mediated by imprinted genes. These studies support the notion that imprinting DMRs could be a vehicle for trans-generational epigenetic inheritance, or at least associated with paternal age.

5.2.5.2 Transposable Elements

Other elements that have been shown to escape complete epigenetic resetting during gametogenesis/embryogenesis include several classes of repetitive elements that are common in the mammalian genome. LINE-1 elements, for example, are highly repeated repetitive elements found throughout the eukaryotic genome (Singer 1982), accounting for approximately 17% of the human genome sequence (Lander, Linton et al. 2001). Approximately 3000–4000 copies remain in a full-length form in the human genome, some of which may be retrotranspositionally active (Sassaman, Dombroski et al. 1997). These elements make a large contribution to diversity across the eukaryotic genome, for example by acting as recombination hotspots and sites for insertions and deletions (Kazazian and Moran 1998; Eickbush and Furano 2002; Deininger, Moran et al. 2003). Recent evidence suggests that transposition of these elements may also occur during later development and adulthood, including in the brain leading to mosaicism (Baillie, Barnett et al. 2011). Human LINE-1 encodes for two proteins (ORF1p and ORF2p) which are required for retrotransposition (Moran, Holmes et al. 1996). LINE-1 elements are generally highly methylated in all tissue types including sperm (Shen, Kondo et al. 2007) and hypomethylation leading to activation of sense and antisense transcripts of these elements has been associated with the development of cancer (Roman-Gomez, Jimenez-Velasco et al. 2005).

IAP elements are another class of retrovirus-like element and are present in over 1000 copies in the mouse genome, which actively transpose via the reverse transcription of an RNA intermediate (Kuff and Lueders 1988; Heidmann and Heidmann 1991). Differential DNA methylation across IAP elements have been found to directly affect gene expression and function, famously exemplified by the Agouti viable mutant mouse in which the insertion of the IAP retrotransposon in the opposite orientation upstream of the agouti gene influences cryptic transcription, resulting from a bidirectional promoter in the 5' long terminal repeat (LTR) of the IAP (Kuff and Lueders 1988). IAP and other transposable elements may need to be

highly methylated in germ line cells because demethylation could cause dysregulation of nearby cellular genes and induce new mutations via replicative transposition (Yoder, Walsh et al. 1997). As these elements are so abundant in the genome and appear to sometimes escape complete epigenetic reprogramming, it is plausible that, like imprinting DMRs, they enable the transmission of epigenetic information transgenerationally. This chapter will therefore also explore the hypothesis that these regions are differentially methylated in the offspring of old compared to young fathers.

5.3 Aims

The aim of this chapter is to investigate epigenetic changes associated with advanced paternal age using the mouse model described in Chapter 3. Using samples from several brain regions and spleen, global DNA methylation will be assessed, in addition to DNA methylation across specific DMRs regulating the expression of brain-expressed imprinted genes, and transposable elements (IAP and LINE-1). Finally, using sperm samples obtained from young, old, and very old males, specific regions associated with advanced paternal age in the offspring will be examined for similar age-associated changes in sperm.

5.4 Methods

5.4.1 Samples

Dissected tissues were obtained from the offspring and breeder mice used in the behavioural analyses presented in Chapter 3, as explained in (section 3.4.8). Sperm samples were collected from an independent collection of C57BL/6J male mice obtained from the MRC Harwell unit, Harwell Science and Innovation Campus. The cauda epididymis is sliced six to eight times with a 30G insulin needle, and then the sperm is gently palpitated out with a pair of forceps. For the vas deferens, forceps are used to “walk” the sperm along the lumen and extrude it through the cut end. Harvested sperm from the cauda and vas were extracted into the lid of a petri dish, after which 50ul of PBS was added and aspirated into an eppendorf before freezing at -80°C. The sperm donors were not used in the breeding of offspring mice used for behavioural or molecular testing.

5.4.2 Brain Tissue, Spleen and Sperm DNA Isolation

DNA from cerebellum and hippocampus were extracted using an optimised phenol chloroform method outlined in detail in the general methods chapter (section 2.2.3). DNA extractions from spleen were performed using the Qiagen DNeasy kit (section 2.2.1), and DNA from frontal cortex was extracted using the Qiagen AllPrep RNA/DNA (section 2.2.2), with the RNA isolated from the same tissues used in the gene expression analyses presented in Chapter 6. Again, full details on these procedures are given in Chapter 2.

DNA from sperm samples was extracted using an adapted phenol-chloroform protocol. 100µl of lysis buffer and 10µl PK at 18.5mg/ml was added to sperm samples. The samples were then incubated overnight (16hrs) at 50^{o}C followed by PK deactivation at 65^{o}C for 20mins. 150µl of PCI was added to the PK treated samples and then mixed by inverting and spun at 13000rpm for 20mins. The aqueous solution was pipetted off into a new eppendorf tube and 150µl of chloroform was added and mixed by inverting then spun at 13000rpm for 5mins. The aqueous solution was again pipetted off into a new eppendorf tube and 150µl of chloroform was added and mixed by inverting then spun at 13000rpm for 5mins. The aqueous solution was pipetted off and 375µl of cold 100% ethanol and 4.8µl of 10M NH_4AC was added and mixed by slowly inverting. The samples were incubated at -20^{o}C for 1hr. After the incubation, the samples were spun at 13,000rpm for 20mins. The supernatant was pipetted off and 400µl 70% ethanol was added followed by a 10min 13000rpm centrifuge. The supernatant was pipetted off and the resulting pellet was allowed to dry at room temperature for 1hr. Finally the samples were resuspended in 100µl DNase/RNase free water. Samples were quantified using NanoDrop NO-1000 using the DNA-50 for concentration and purity (260/280 greater than 1.8) and checked on a 0.8% agarose gel for DNA integrity.

5.4.3 Luminometric Methylation Assay

The LUminometric Methylation Assay (LUMA) was used to estimate global DNA methylation by interrogating all CCGG recognition sites across the genome (Karimi, Johansson et al. 2006; Karimi, Johansson et al. 2006). The method is based on the fact that both HpaII and MspI cut DNA at the recognition site CCGG but because HpaII is a methylation sensitive restriction enzyme (MSRE) it only cuts when DNA is unmethylated. MspI, on the other hand, cuts regardless of DNA methylation status. LUMA was run on DNA from spleen from both the

offspring and fathers, on DNA from cerebellum and hippocampus from the offspring, and in sperm samples from unrelated males of different ages. In this protocol, high molecular weight DNA was diluted to 500ng in 17µl of DNase free water so the final concentration was 29.4ng/µl. All samples were processed in duplicate for two separate digestions, and all experiments included a fully methylated and fully unmethylated control to calibrate the data and control for incomplete digestion. The enzyme mixes were prepared as below. 3µl of the Mix A was added to each of two 17µl DNA samples, and 3µl of Mix B was added to another two aliquots. The samples were then incubated at 37^{o}C for 4hours (*Figure 5.2*). EcoRI was included in all reactions as a normalization reference.

HpaII - Mix A		*MspI - Mix B*	
Tango Buffer	2µl	Tango Buffer	2µl
EcoRI (10U/µl)	0.5µl	EcoRI (10U/µl)	0.5µl
HpaII (10U/µl)	0.5µl	MspI (10U/µl)	0.5µl

After incubation, 20µl of Pyrosequencing annealing buffer was added to each tube. The samples were then run on the Qiagen Q24 pyrosequencer. Since different overhangs are produced after DNA cleavage, i.e. both MspI and HpaII have 5‘-CG overhangs and EcoRI has 5‘-AATT overhangs, they can be filled in by separate polymerase extensions. In the LUMA assay, dNTPs are added in four sequential steps: First dATPS, then the mixed dGTP + dCTP followed by dTTP and a final dGTP + dCTP. The incorporation of a dNTP to the sequence was accompanied by the release of pyrophosphate (PPi) in a quantity of moles equal to the amount of the incorporated nucleotide. PPi was converted to ATP in the presence of 5' phosphosulfate (APS) and the ATP then drives the conversion of luciferin to oxyluciferin that generates light in an amount proportional to the amount of ATP created. This light is detected by a charge coupled device (CCD) chip and seen as a peak in the raw data output. The height of the peak generated by the light signal is proportional to the number of nucleotides added. Unincorporated nucleotides and ATP were degraded after each dNTP adding step and when the degradation is complete, the next addition event occurs. The light peaks produced by the addition of both dATPS and dTTP represents the EcoRI cleavage and are therefore expected to be equal to in both enzyme digests and are used as an internal control for the assay. The peak produced by the addition of the first dGTP and dCTP represents HpaII or MspI cleavage. The

second addition of dGTP and dCTP is used as a control to signify the completion of the first dGTP and dCTP addition and is expected to be close to zero.

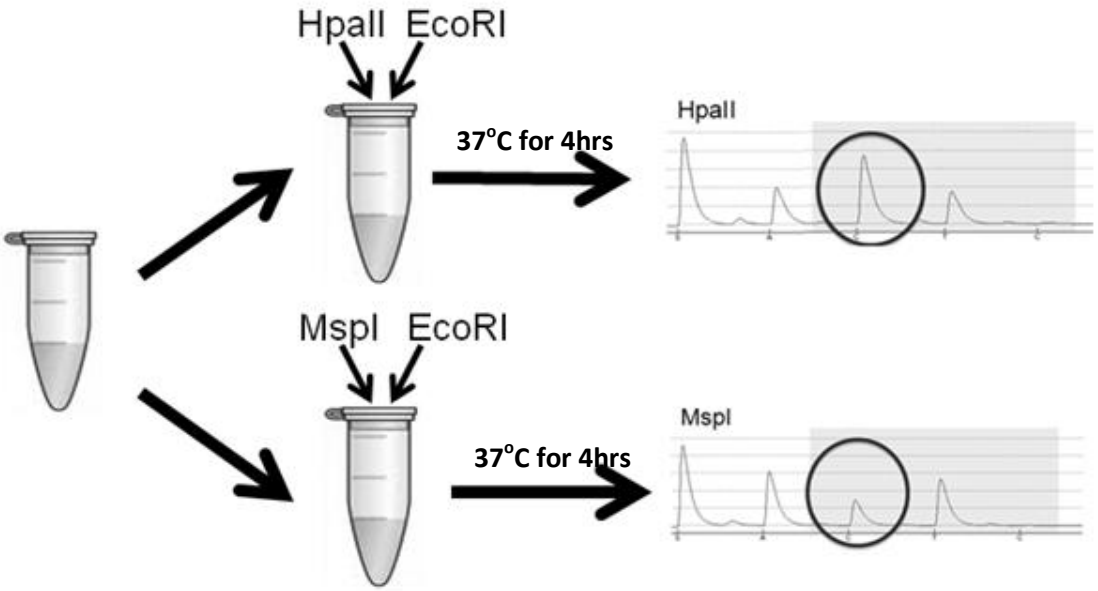


Figure 5.2 - LUMA Method

DNA sample split and one aliquot incubated with HpaII and EcoRI and one aliquot incubated with MspI and EcoRI. The cut samples are then analysed by pyrosequencing

5.4.4 LUMA Data Analysis

After the peak heights are exported from the Pyrosequencer, the ratio of the C (HpaII or MspI overhang) peak is divided by the A peak (EcoRI overhang) for both HpaII and MspI separately (*Figure 5.3*). This is repeated for the duplicate sample and then the average of C/A is taken. The inclusion of EcoRI as an internal control provides an accurate HpaII/MspI ratio for each sample of which the HpaII/MspI ratio can be defined as (HpaII/EcoRI) / (MspI/EcoRI). The average value of HpaII ratio is then divided by the average MspI Ratio. The fully methylated control should be low (we generally observe a value around 0.02) and the unmethylated control is normally much higher (around 1.22).

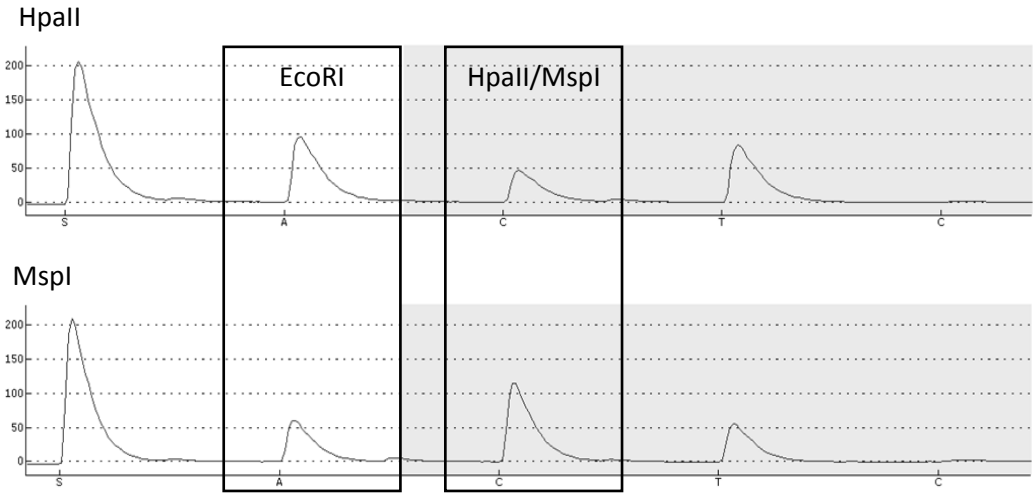


Figure 5.3 – Example Luminometric Outputs from the LUMA Assay

Top panel shows output from HpaII and EcoRI digest and the bottom panel shows the output from the MspI and EcoRI digest from the same individual. The 'S' peak is the substrate peak, 'A' peak is from the overhang from EcoRI digest, the first 'C' is the C+G addition for the HpaII or MspI overhang. Methylation is calculated by dividing the height of the first 'C' peak by the height of the 'A' peak for each digest then dividing the HpaII (C/A) by MspI (C/A). For the output above the values would be HpaII 94.83/47.54 = 0.50, MspI 60.56/115.05 = 1.90, HpaII/MspI = 0.26.

5.4.5 Sodium Bisulfite Treatment

Sodium bisulfite treatment of each DNA sample was carried out using an optimised protocol as outlined in the Material and Methods chapter (section 2.3).

5.4.6 Sequenom EpiTYPER Analysis

A number of methods exist for the quantitative assessment of DNA methylation at specific CpG sites (Laird 2010). In this project, we used the Sequenom EpiTYPER platform which provides a method for accurately measuring methylation at CpG units based on cleavage at thymine residues combined with Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) (MALDI-TOF) mass spectrometry (Sequenom 2011) (*Figure 5.4*). Bisulfite PCR products are tagged with a T7 tag on their reverse strand which acts as a primer for reverse transcription. Uracil-specific cleavage creates fragments for analysis in the mass spectrometer

and the resulting cleavage pattern depends on the presence of methylated cytosine in the original genomic DNA. Although Sequenom EpiTYPER is one of many ways of quantitatively assessing DNA methylation across specific sites, it is relatively high throughput, quick and can provide quantitatively accurate results at multiple CpG dinucleotides within an amplicon.

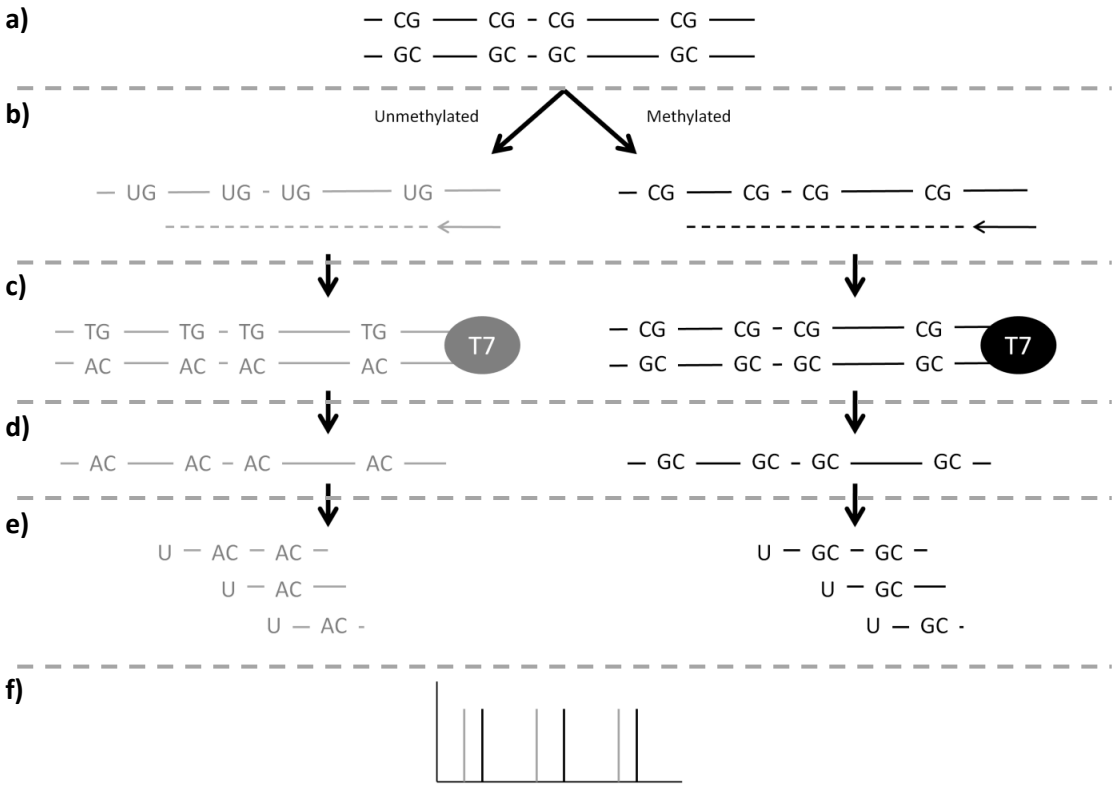


Figure 5.4 - Sequenom EpiTYPER Methodology

Adapted from Sequenom EpiTYPER protocol (Sequenom 2011)

a) *Genomic DNA*
b) *Bisulfite treatment of DNA (section 2.3)*
c) *PCR and SAP treatment – Amplifies up the region of interest and incorporated as T7 tag to each amplicon and SAP treatment removed unincorporated dNTPs*
d) *In vitro transcription - Carried out on the reverse strand of amplicons*
e) *Uracil-specific cleavage – Base specific cleavage to create fragments for analysis on mass spectrometer*
f) *MALDI-TOF Mass Spectrometry – Size of the fragments are measured by mass spectrometry and ratio of peak heights related to CpG units are converted into percentage methylation*

5.4.6.1 Sequenom EpiTYPER Primer Design

PCR primers were designed to amplify specific target regions of interest using the Sequenom EpiDesigner online software (www.epidesigner.com) (Sequenom 2007). Bioinformatic analyses were used to obtain the genomic sequence around areas of interest, focussing on CpG islands known to be imprinting DMRs. The following general primer design parameters were used:

	Min	Optimal	Max
Primer Tm (^{o}C)	56	62	64
Primer Size (bp)	20	25	30
Product Size (bp)	100	200	300

Primers were selected to be designed on either the forward or reverse strands, contain at least five CpG sites and to analyse only T cleave reactions. It was also important that primer sequences contained no CG dinucleotides and no SNPs to ensure efficient and non-biased amplification. Where several primer-pair options were available, the set spanning the highest number of CpGs was selected. As the sequenom assay is based on base specific cleavage, some fragments may contain more than one CpG site. DNA methylation estimates are obtained for individual 'CpG units' which can contain one or multiple CpG sites. All the designed primers for the selected imprinting DMRs are listed in *Table 5.1* (for images showing the location of the imprinting DMRs assessed, see *Appendix 2*). The assay names used in this chapter relate to the imprinted gene each specific DMR is associated with. Where more than one gene is regulated by a given DMR, as is often the case in imprinted gene clusters, only one name to identify the amplicon location, but other genes affected by the DMRs will be discussed in the relevant sections.

Bisulfite-PCR primers were initially tested using a standard PCR cycling protocol (section 2.4) (annealing temperature of 55^{o}C for 35 cycles with 0.5µl of $MgCl_2$). PCR primers were tested on pooled bisulfite treated mouse test samples to optimise assays. Products were visualised using agarose gel electrophoresis (section 2.5), and where necessary (i.e. where there was no amplification, incorrectly-sized PCR products or non-specific amplification, etc), were further optimised using changes to annealing temperature or adjusting the amount of $MgCl_2$ added to the reaction.

Table 5.1 - Imprinting DMR Primer Details

** A tag of AGGAAGAGAG is added to the 5' end, ** A tag of CAGTAATACGACTCACTATAGGGAGAAGGCT is added to the 5' end*

Assay Name	Genomic Position (mm9)	Expressed Allele	Methylated Allele	Location	Left Primer*	Right Primer**	Optimised Annealing Temp (°C)	PCR Length (bps)	Fragments (CpGs Covered)
Cd81	chr7:150238650-150238935	Maternal	Paternal	Promoter	TTTAGAGGTTTATAAAGAGTGAGGAGT	AATACATTTAATACAACCCTCCACC	55	285	25 (27)
Exon1A	chr2:174154376-174154704	Paternal	Maternal	Gnas Promoter/DMR	GAAGAAGAAGAAGAATGTGTTAGAAGTTT	CAAAAAATTAAACACCCATAAAAAA	55	328	25 (36)
Grb10	chr11:11937789-11937969	Isoform Dependent	Maternal	Upstream CpG island	GTATAGTTTAAATTTTTTTTGGGTTG	ATAAAAAACCTCTCAAACTTCCCTC	55	180	12 (15)
Gtl2	chr12:110766411-110766776	Maternal	Paternal	Upstream CpG island	TGTTGTGGTTTGTTATGGGTAAGTT	ATCCCCTATACTCAAAACATTCTCC	55	365	28 (29)
Igf2	chr7:149846224-149846584	Paternal	Maternal	Promoter	GTTTTTGGGATTTTATTTAGTATGAG	CTTAAAACCTCCAACCCTACAAATC	55	360	17 (17)
Inpp5f	chr7:135831573-135831807	Paternal	Maternal	Promoter Short form	TGGGTAGAGGGTTGTTAAGTTGTTA	ACCCTCCTCAACTATATATCCTTCC	55	234	11 (11)
Kcnq1ot1	chr7:150481246-150481473	Paternal	Maternal	Intron 11	GTATTTTATTTATTATTTTGGTGTTGGTT	CCCCATCTTTATAACCCAAACTTAC	55	227	10 (11)
Mcts2	chr2:152512486-152512694	Paternal	Maternal	Promoter Mcts2/H13	GAGTTGAGGGATGGATAGTTATTT	CAAAATTCCCAAACATTAAAATACTAAAA	55	208	15 (17)
Mest	chr6:30687833-30688178	Paternal	Maternal	Promoter	TTTTATTAGAATTTGGGGTTTAGGA	CAACAAAAACAACAAACAACAACTC	55	345	19 (22)
Nap1l5	chr6:58856789-58857051	Paternal	Maternal	Promoter	TTATAAAGTTTTTTTTGGGTTTGGG	CTACAAAACCTCTCTAAACCAACTCT	55	262	9 (23)
Gnas-Nesp	chr2:174110117-174110436	Maternal	Paternal	Gnas Promoter/DMR	GGGTTATTTTTTGTAGAGTTAGAGGG	CAAACTCAAATTCAAAAAACTCAAAA	55	319	23 (25)
Nnat	chr2:157386018-157386034	Paternal	Maternal	Promoter	GTTTTGGGTGGGAGAGGGTATT	AACTATAAAATTTCACAACACACAAA	55	336	23 (32)

Sgce	chr6:4697698-4697923	Paternal	Maternal	Promoter	AAATTTTGTTAAGTTTTTAGTGGTTAGA	CTTAAAAATACAAAACCAATCACTTT	55	225	11 (16)
Snrpn	chr7:67149900-67150145	Paternal	Maternal	Intronic/ Snurf Promoter	TTTTGGTAGTTGTTTTTTGGTAGGA	AAATCCACAAACCCAACTAACCTTC	55	245	11 (14)
Ube3a	chr7:66483985-66484228	Maternal	Paternal	Promoter	GAGTTAGAGGAGGGTTTTATAGAGTTG	TCTCCTAAAATCACCCAAAAAAAA	59	243	21 (27)
XL	chr2:174124857-174125236	Paternal	Maternal	Gnas Promoter/DMR	GGGTAGTTTTTGTTATTTAGGTGGA	CTAAAAAATCTAAAATAAACCCTACTAAAA	57	379	23 (23)
Zac1 (1)	chr10:12810944-12811233	Paternal	Maternal	Plagl1 Promoter	TTTTGAATTTGGGTGTTTTAGTTGT	AAATCCAACATTTTCAAATAACTTT	55	289	12 (15)
Zac1 (2)	chr10:12810308-12810540	Paternal	Maternal	Plagl1 Promoter	GTTAGTTAAGGGGATGGTTGTATGG	CAAATTTTCAAACTAAATTAAAACTACC	55	232	10 (15)
Zim2	chr7:6682272-6682554	Maternal	Paternal	Peg3 Intron 1	TTTGAGTAATTTTTTTTGTGGTTTT	TATACAACAAACCTATCTCCACCCT	55	282	13 (15)

Primers for transposable elements are listed in *Table 5.2*. The primer details for both the IAP and LINE-1 assays were designed by and obtained from Prof. Wolf Reik's group (Babraham Institute, University of Cambridge) and used previously in their published studies of these elements (Popp, Dean et al. 2010). Roughly 1000 copies of IAP are present in the mouse genome, and all are very similar in sequence regardless of their genomic location. Due to this the IAP Sequenom assays provide a method for assaying the majority of these sites and providing an average DNA methylation level. LINE-1 elements are also present in many copies, and the LINE-1 Sequenom assay is again designed to give an average methylation level across multiple LINE-1 elements throughout the genome (*Figure 5.5*).

Table 5.2 - Transposable Element Primer Details

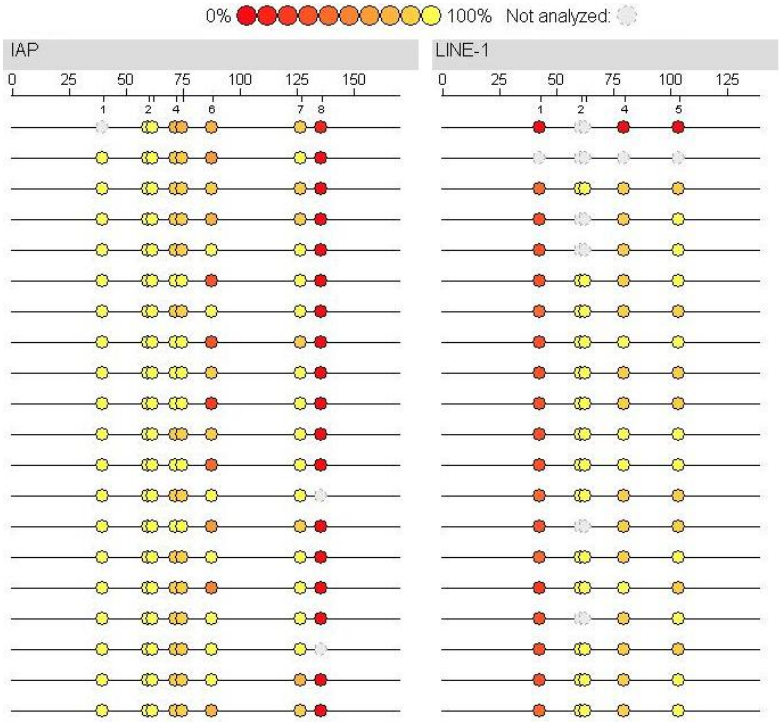
** A tag of AGGAAGAGAG is added to the 5' end,*

*** A tag of CAGTAATACGACTCACTATAGGGAGAAGGCT is added to the 5' end*

Name	Forward Primer *	Reverse Primer **	Annealing Temp (°C)	PCR Length (bps)	Fragments (CpGs Covered)
IAP	ATGGGTTGTAGTTAATTAGGGAGTG	ACACATAAAATAAAAAACCACCCTC	55	150	6 (8)
LINE-1	GATTTTAAGATTTTTGGTGAGTGGA	AAAAACTTATACCCCAAATCAAACC	55	119	4 (5)

Figure 5.5 – Example Epigram of IAP and LINE-1 Sequenom Assay

Individual rows relate to samples and columns are CpG units covered in the amplicon. Red circles indicate low levels of methylation and change on a gradient towards yellow, which indicates high methylation. Empty circles are CpG units not analysed.

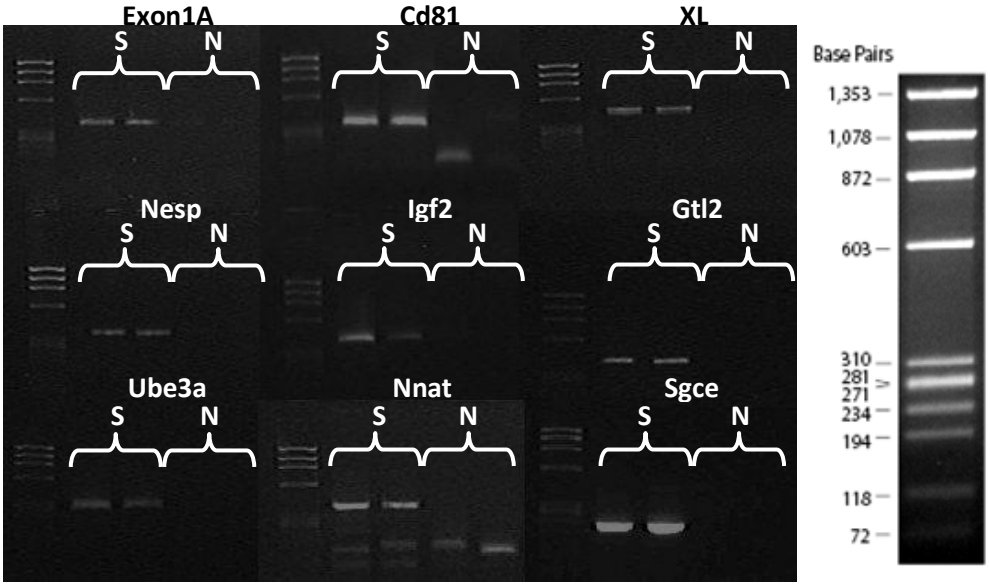


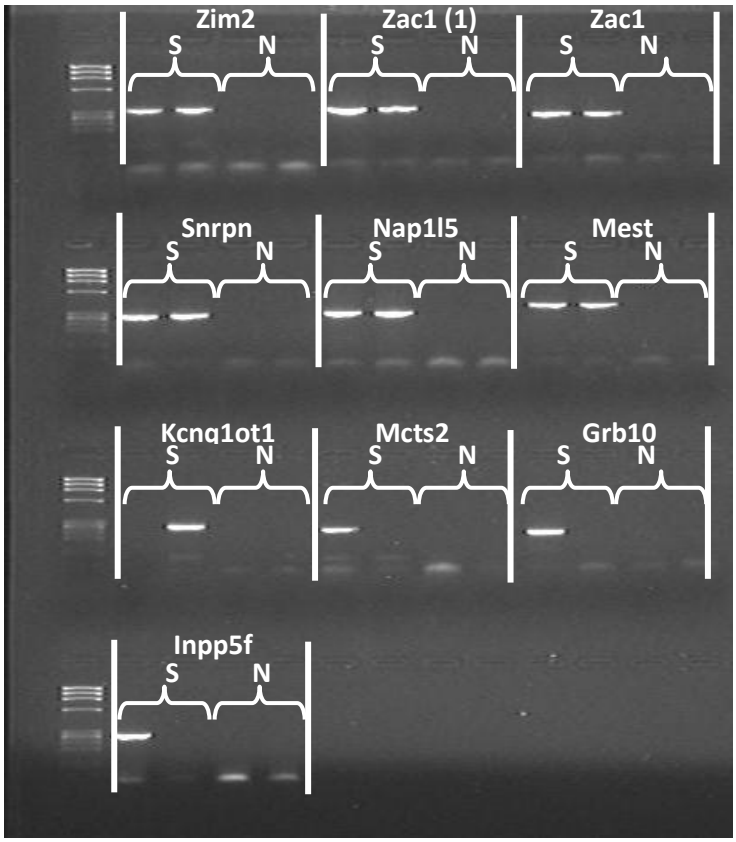
5.4.6.2 Polymerase Chain Reaction

All PCRs were carried out on bisulfite converted DNA in duplicate using the protocol outlined in the general methods chapter (section 2.4) using the optimised annealing temperatures. PCR products were run on 1.5% agarose gels as described in the general methods chapter (section 2.5) to check specificity of PCR primers and correct size of bands (*Figure 5.6*).

Figure 5.6 - PCR Amplicons

S – Sample, N – Negative Control. All PCR products checked on a 1.5% agarose gel against a ΦX174 DNA-HaeIII ladder (pictured below). All PCR were carried out with a fully methylated and fully unmethylated control and two negative controls.





5.4.6.3 Sequenom EpiTYPER Protocol

After checking PCR amplification by gel electrophoresis, unincorporated dNTPs were removed using SAP cleanup. The SAP enzyme solution was made according to the recipe below and 2µl of the SAP mix added to each well, which was then incubated on a thermocycler at 37°C for 40mins before the SAP was denatured at 85°C for 5mins.

SAP Mix

RNase free water	1.7µl
Shrimp alkaline phosphatase (SAP)	0.3µl

After SAP treatment, the hMC transcription/RnaseA mix was made as below and then 5µl of the mix was added to each well. The samples were then incubated at 37°C for 3 hours.

hMC transcription/RnaseA Mix

DNase free water	3.21µl
5X T7 Polymerase buffer	0.89µl
Cleavage mix (T mix)	0.22µl
DTT (100mM)	0.22µl
T7 R&DNA Polymerase (50U/µl)	0.4µl
RNase A	0.06µl

20µl of DNase free water was then added to each well and the plate spun down. The addition of resin was used to desalt the samples before being spotted on the SpectroCHIP and scanned. Using the 384 well 6mg Dimple Plate and scraper, each of the dimples on the plate was filled with resin by placing a large amount at the left side of the plate and scraping the resin down the plate making sure all the wells were properly filled (*Figure 5.7*). The plate was then left for 15-20mins to dry.

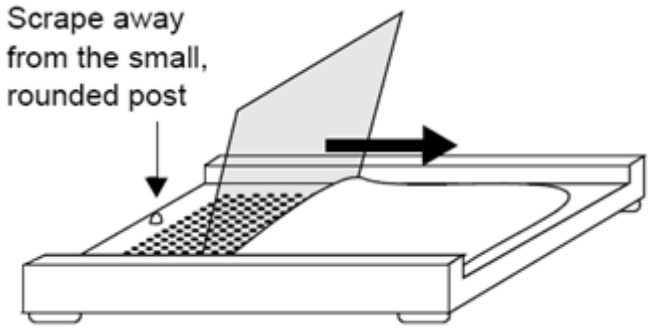


Figure 5.7 - Resin Addition to 6mg Dimple Plate

Resin addition is controlled by plating out in the 6mg dimple plate. The resin is added to each of the samples and the plate it rotated for 15mins (Sequenom 2011).

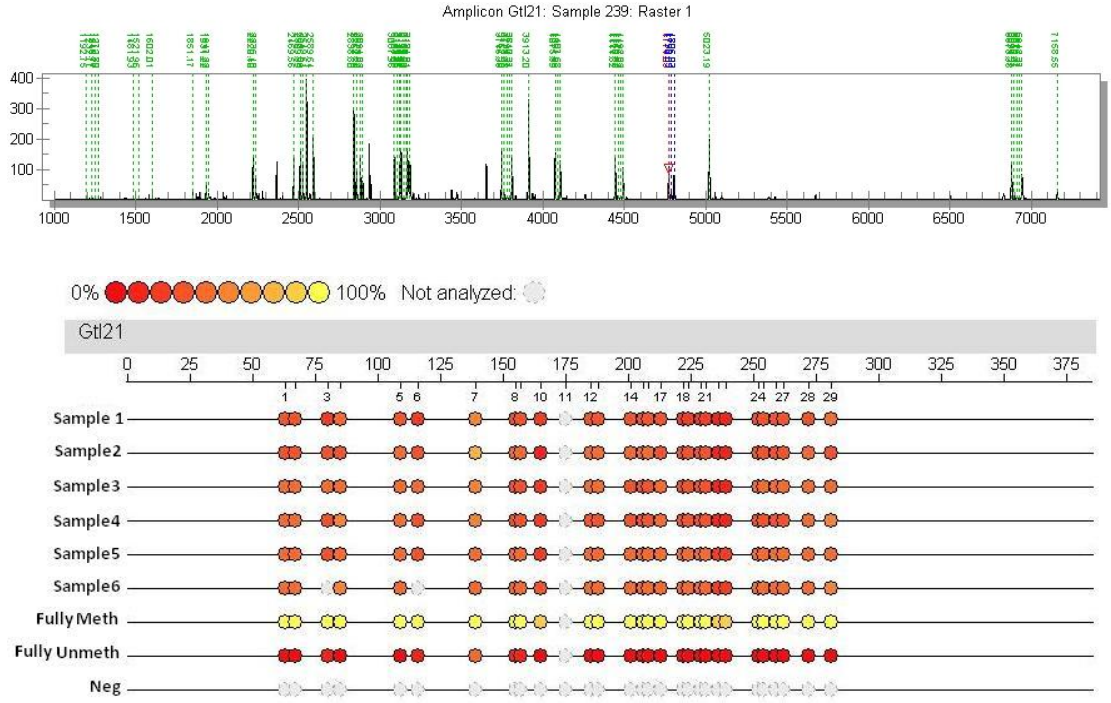
When the resin was dry in the dimple plate the 384 well sample plate was placed upside down on the top of the dimple plate. The dimple plate and sample plate are turned over together so all the resin fell into the wells of the sample plate. The plate was then rotated slowly for 15mins on plate rotator at a slow speed. The plate was then centrifuged for 5mins at 3000rpm. The samples were then spotted on to the SpecroCHIP and scanned in the Sequenom MassARRAY. Positive controls, including both artificially methylated and artificially unmethylated samples were included in all experimental procedures.

5.4.6.4 Sequenom EpiTYPER Data Analysis

After the fragment mass is determined by MALDI-TOF MS, the EpiTYPER software generates a report that contains quantitative information for each analyzed fragment. All negative controls were checked for empty epigrams and fully methylated and fully unmethylated controls checked for their corresponding methylation levels (*Figure 5.8*). CpG units with low call rates (<80%) were removed from subsequent analysis, as were samples with less than 70% data available across the CpG unit. All data from CpG-containing fragments flagged by EpiTYPER as having low mass (outside MS analytical window), high mass (outside MS analytical window) or silent peak overlap (two overlapping peaks, one with no CpGs, following MS) were discarded.

Figure 5.8 – Example Spectra and Epigram from EpiTYPER Software

Spectral reading and epigram output from Sequenom EpiTYPER software showing examples of fully methylated controls, fully unmethylated controls and negative controls.



Differences between groups were calculated using the student's t-test and one-way ANOVA in the R statistical analysis environment. A 'between-within' analysis was used to test DNA changes in the DNA methylation pattern across the PCR-amplified regions, with parental age as the between factor and CpG site as the within factor. Significance level for all tests was set at 0.05.

5.5 Results

In all graphs shown in this section, ** = $p \leq 0.01$, * $p = \leq 0.05$ and + = $p \leq 0.1$. Error bar in all graphs show the SEM.

5.5.1 Global DNA Methylation Analysis using LUMA

In this section global DNA methylation was assayed in multiple tissues from the offspring of young fathers (n = 18) and the offspring of old fathers (n = 23). We also assessed DNA obtained from the offspring of very old fathers (n = 11), although these samples were obtained from a different batch of animals than the other two groups, and thus the data for these may not be directly comparable (as described in section 3.4.1). Furthermore, the offspring of 'very old' fathers analysis group was considerably smaller than the other groups, and so they were also combined with the offspring of old fathers into a single analysis group consisting of the offspring of fathers over 10 months (n = 34). The HpaII/MspI ratio is inversely related to DNA methylation level, therefore a low HpaII/MspI ratio corresponds to a high methylation level and vice versa.

5.5.1.1 Spleen

The LUMA data from the spleen DNA from the offspring is shown in *Figure 5.9*a. DNA methylation was seen to increase with the age of the father, with the offspring of very old fathers being significantly more methylated than the offspring of young fathers ($t = 3.95$, d.f. = 20, $p = 0.002$) and the offspring of old fathers ($t = 3.58$, d.f. = 23, $p = 0.002$). Combining the offspring of old fathers and the offspring of very old fathers (i.e. the offspring of fathers over 10 months old) again highlighted significant hypermethylation in the offspring of older fathers compared to the offspring of young fathers ($t = 1.94$, d.f. = 34, $p = 0.05$) (*Figure 5.9*b).

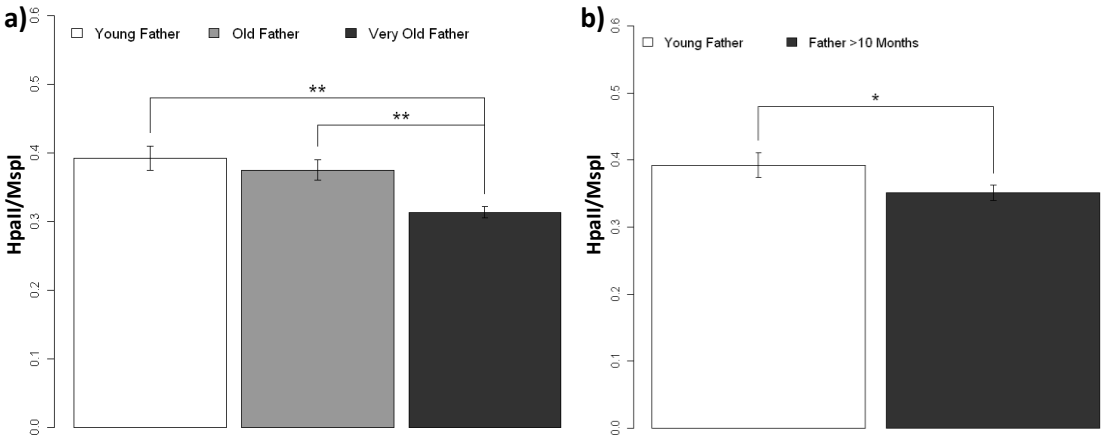


Figure 5.9 - Results from LUMA on Offspring Spleen DNA

a) Offspring of young fathers n=13, Offspring of old fathers n=14, Offspring of very old fathers n=9.

b) Offspring of young fathers n=13, Offspring of fathers>10months n=23

When global methylation was assayed in spleen samples from the fathers themselves, methylation levels were not significantly different between groups ($F(2, 12) = 0.38$, $p = 0.69$) (*Figure 5.10*). This data suggest that the association between global hypermethylation and advanced paternal age detected in the spleen of offspring are not reflected in somatic spleen cells from the fathers.

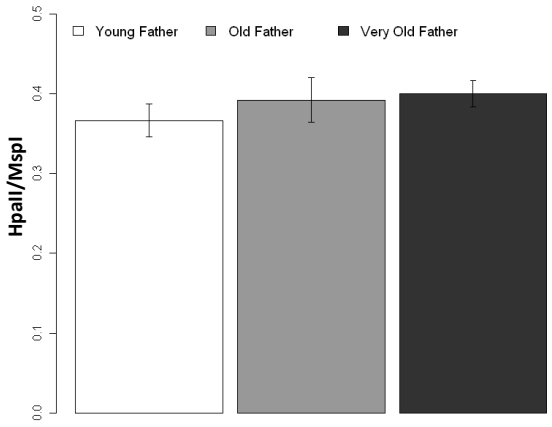


Figure 5.10 - Results from LUMA on Fathers Spleen DNA

Young fathers n=4, Old fathers n=6, Very old fathers n=3

5.5.1.2 Cerebellum

We next attempted to replicate this observation in an additional tissue from the same animals, the cerebellum. Although DNA methylation levels between the offspring of young fathers and the offspring of older fathers did not reach statistical significance, a near-significant trend in the same direction in global methylation was observed as in the spleen (t = 1.77, d.f. = 25, p = 0.08) (*Figure 5.11*a), with the offspring of old fathers having a higher level of global methylation. A similar trend towards significance in the same direction is observed comparing the offspring of young fathers with the offspring of fathers over 10 months (t = 1.71, d.f. = 34, p = 0.08) (*Figure 5.11*b)

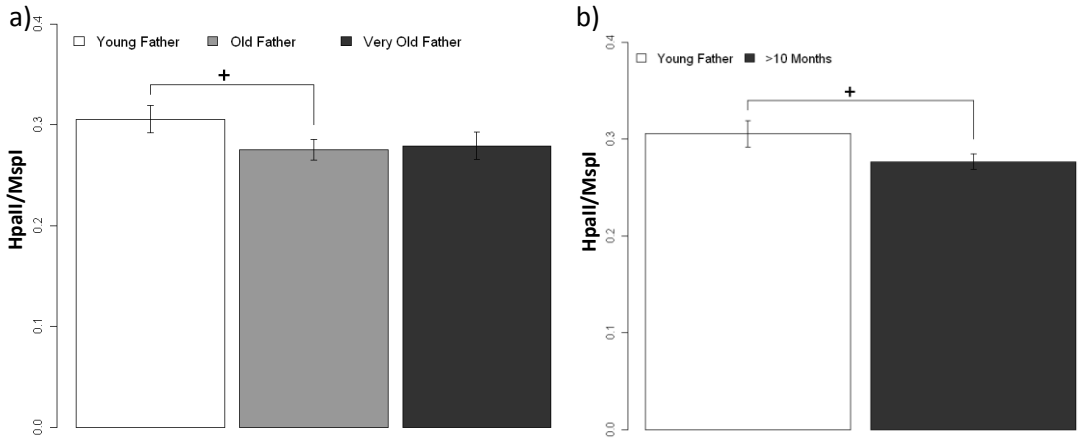


Figure 5.11 - Results from LUMA on Offspring Cerebellum DNA

a) *Offspring of young fathers n=12, Offspring of old fathers n=15, Offspring of very old fathers n=9*
b) *Offspring of young fathers n=12, Offspring of fathers >10months n=24*

5.5.1.3 Hippocampus

In contrast, no such paternal-age associated difference in DNA methylation was observed in hippocampus samples from the offspring, where the global DNA methylation levels between groups were not significantly different either when looking at three groups individually (F (2, 50) = 0.14, p = 0.86) (*Figure 5.12*a), or when combining the offspring of old fathers and very old fathers (t = 0.52, d.f. = 50, p = 0.61) (*Figure 5.12*b). Global methylation in the offspring of old fathers was not significantly different than in the offspring of young males (t = 0.40, d.f. = 32, p = 0.69).

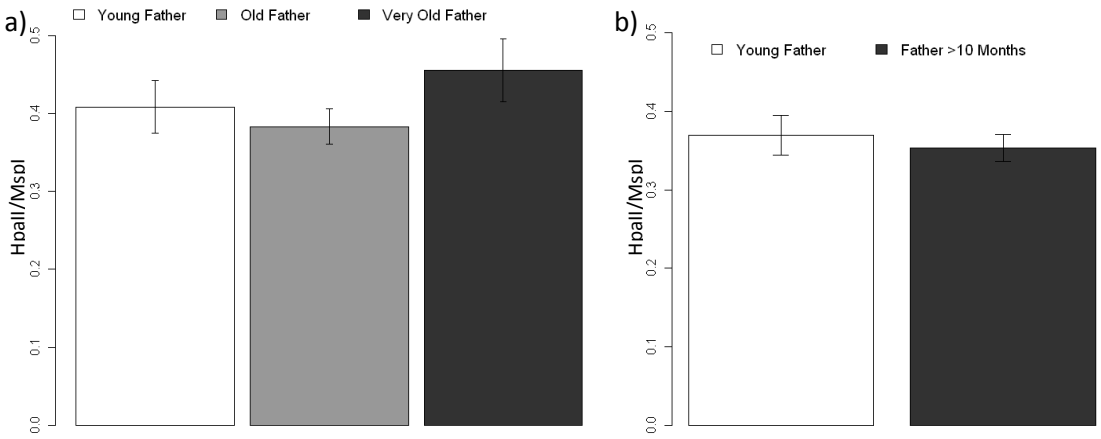


Figure 5.12 - Results from LUMA on Offspring Hippocampus DNA

a) *Offspring of young fathers n=18, Offspring of old fathers n=23, Offspring of very old fathers n=11*
b) *Offspring of young fathers n=12, Offspring of fathers >10months n=24*

5.5.1.4 Sperm

Global methylation was assayed in the sperm from males of different ages unrelated to the mice used in the rest of the study. No differences in global methylation were seen between the young, old and very old males (F (2, 22) = 0.005, p = 0.99) (*Figure 5.13*a), or between the young males and males over 10 months (t = -0.06, d.f. = 22, p-value = 0.95) (*Figure 5.13*b).

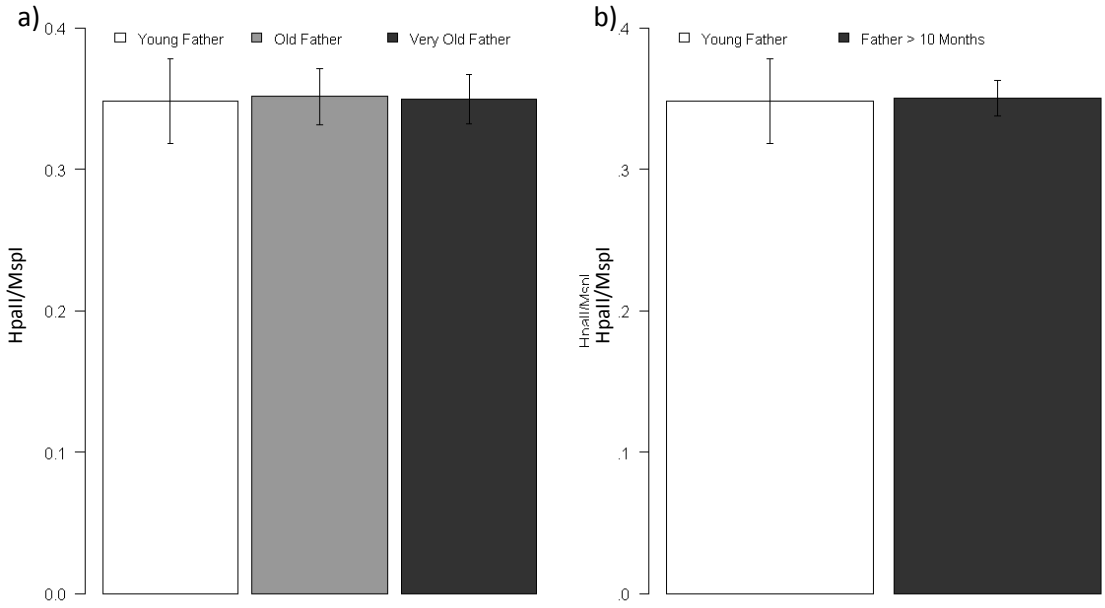


Figure 5.13 - Results from LUMA on Sperm DNA

a) Young males n=6, old males n=10, very old males n=8

b) Young males n=6, males >10months n=18

Figure 5.14 shows the average HpaII/MspI value from LUMA across all tissues assayed. DNA from cerebellum has the lowest HpaII/MspI value and so the highest level of global methylation. Global methylation levels are similar in hippocampus, spleen and sperm samples. The global methylation in cerebellum was significantly higher (lower HpaII/MspI) than in the spleen (t = -6.35, d.f. = 102, p = 2.65E-08), hippocampus (t = -4.55, d.f. = 102, p = 2.14E-05) and sperm (t = -4.52, d.f. = 74, p = 6.90E-05). There were no significant differences between global methylation in hippocampus, spleen or sperm (F (2,126) = 0.26, p = 0.77).

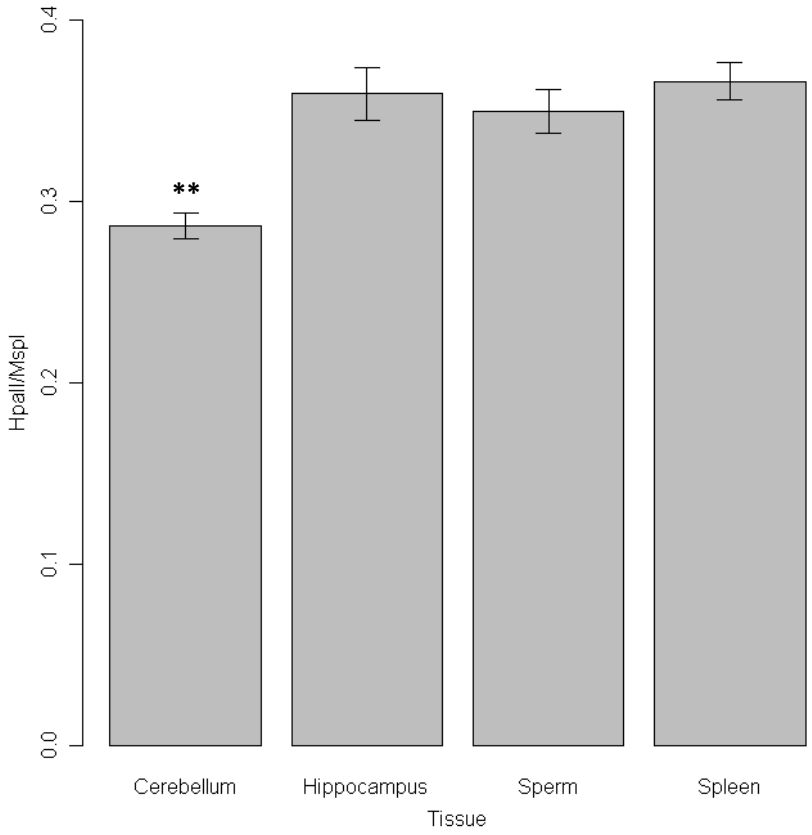


Figure 5.14 – Average HpaII/MspI in All Assayed Tissues from LUMA

Average HpaII/MspI across all groups in all tissue assayed using LUMA combined with pyrosequencing.

5.5.2 DNA Methylation at DMRs of Brain-Expressed Imprinted Genes

In this section, DNA methylation levels across DMRs associated with brain-expressed imprinted genes were assayed in tissues from the offspring of young fathers (n = 18), the offspring of old fathers (n = 23), and the offspring of very old fathers (n = 11). As before, the offspring of old fathers and offspring of very old fathers were also combined (n = 34). In addition, sperm DNA samples from male mice of two (young), ten (old) and twelve (very old) months old were also assessed at several loci. These mice, although from the same inbred mouse strain (i.e. C57BL/6), were not directly related to our test mice.

5.5.2.1 Cerebellum

Again, our initial experiments were carried out on cerebellum samples, in which we assessed DMRs and CpG islands associated with 15 known brain-expressed imprinted loci. The cerebellum was selected as the tissue for our primary analyses due to its postulated role in the development of autism, one of the neuropsychiatric disorders strongly linked to advanced paternal age (Martin and Albers 1995; Schmahmann 2000; Sparks, Friedman et al. 2002; Brambilla, Hardan et al. 2003). As expected for DMRs associated with the monoallelic expression of imprinted genes, most regions were found to be hemi-methylated (~50%) (*Figure 5.15*). Assays associated with Cd81, Ube3a and Igf2 showed much lower average methylation, and so were not run across the entire sample as they did not appear to be DMRs.

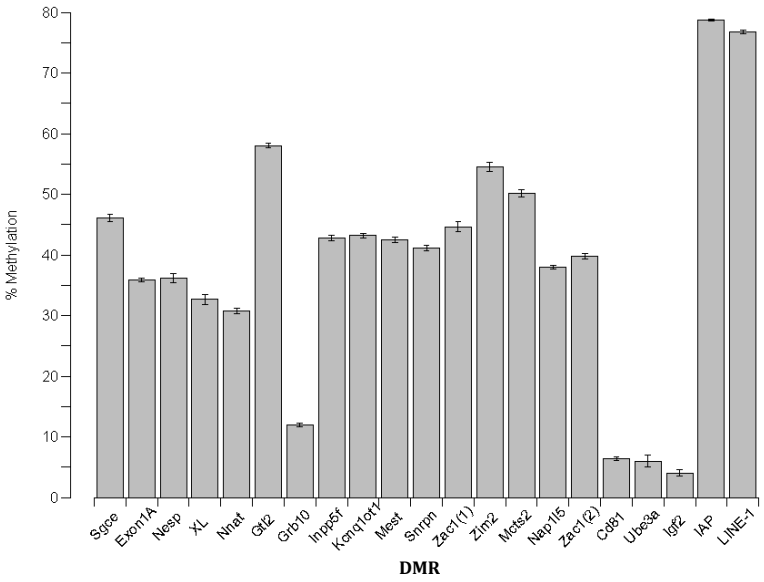


Figure 5.15 –Average Methylation across All Sequenom Assays

Average methylation across all amplicons and all cerebellum samples run on Sequenom EpiTYPER.

First, looking at amplicon averaged levels, there was a modest (~3%) but significant decrease in DNA methylation in the offspring of old fathers compared to the offspring of young fathers in two DMRs: Kcnq1ot1 ($t = 2.47$, d.f. = 39, $p = 0.02$) and Mcts2 ($t = 2.09$, d.f. = 39, $p = 0.05$). Furthermore, a non-significant trend for amplicon-averaged hypermethylation in

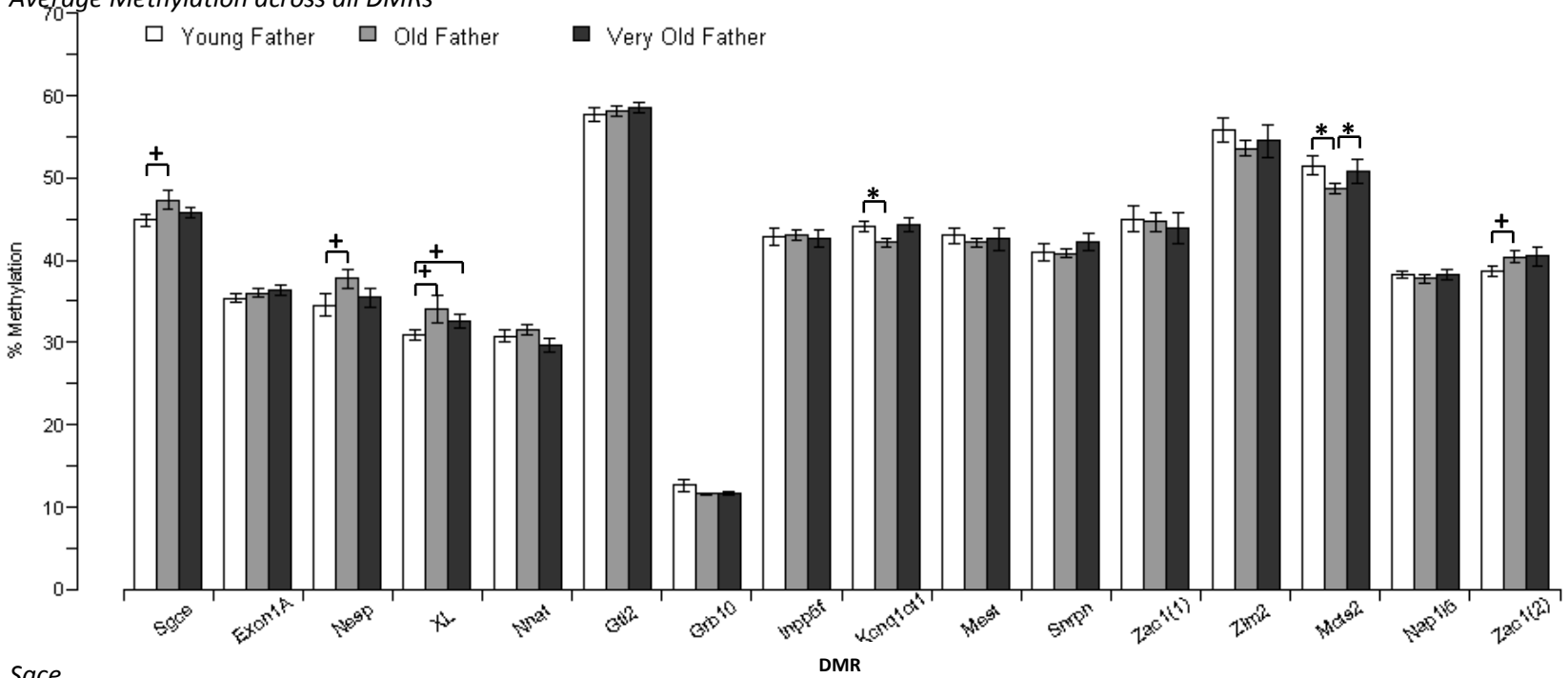
the offspring of old versus young fathers was observed for three DMRs: Gnas-Nesp (t = -1.84, d.f. = 39, p = 0.07); Sgce (t = -1.72, d.f. = 39, p = 0.09) and Zac1 (CpG island 2) (t = -1.90, d.f. = 39, p = 0.06). In addition, many individual CpG sites across the 16 DMRs had significant (or trends towards significant) differences between groups defined by paternal age as summarised in *Figure 5.16* and *Table 5.3*.

Figure 5.16 - Methylation of Imprinted Genes in Cerebellum samples

Average methylation in all DMRs and methylation levels across 16 imprinted DMRs in offspring of young fathers (n = 18), offspring of old fathers (n = 23) and offspring of very old father (n = 11)

Figure 5.16 Continued

Average Methylation across all DMRs



Sgce

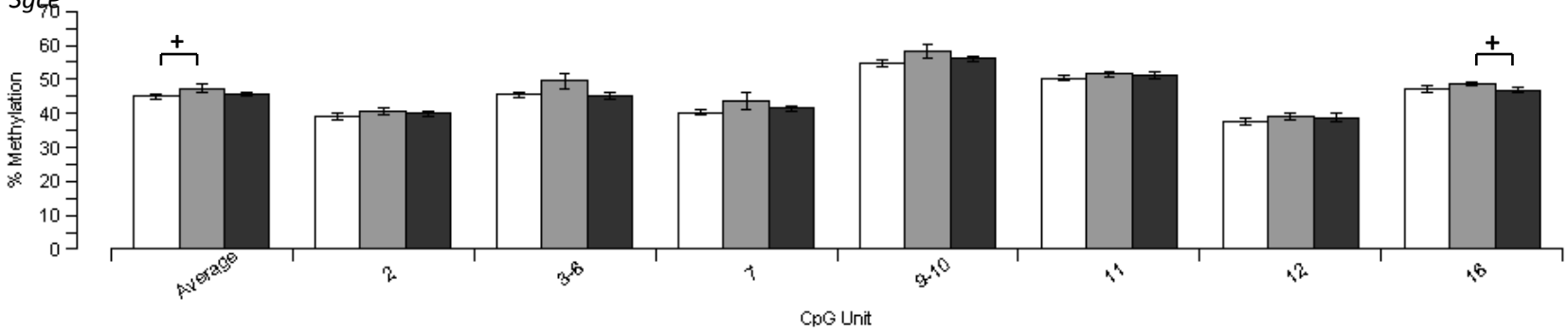
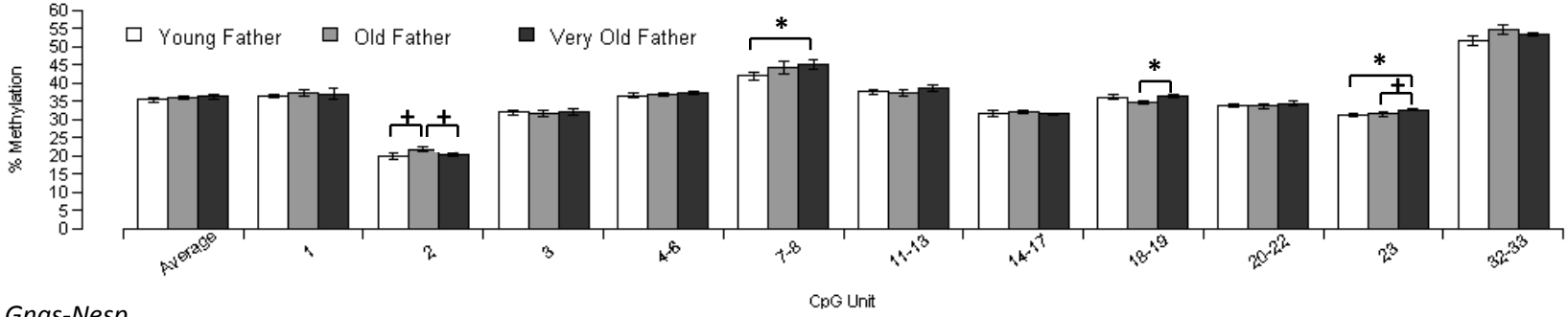
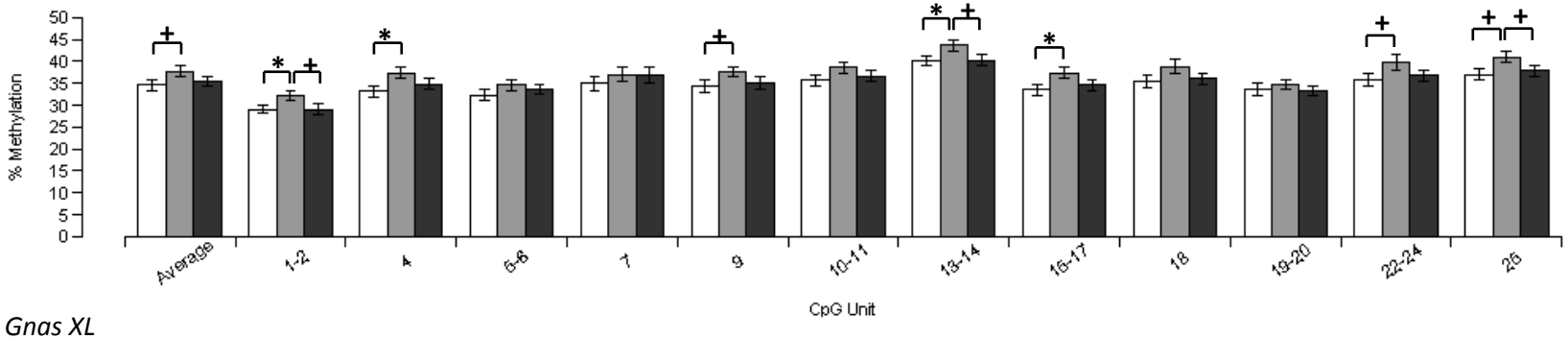


Figure 5.16 Continued

Gnas Exon1A



Gnas-Nesp



Gnas XL

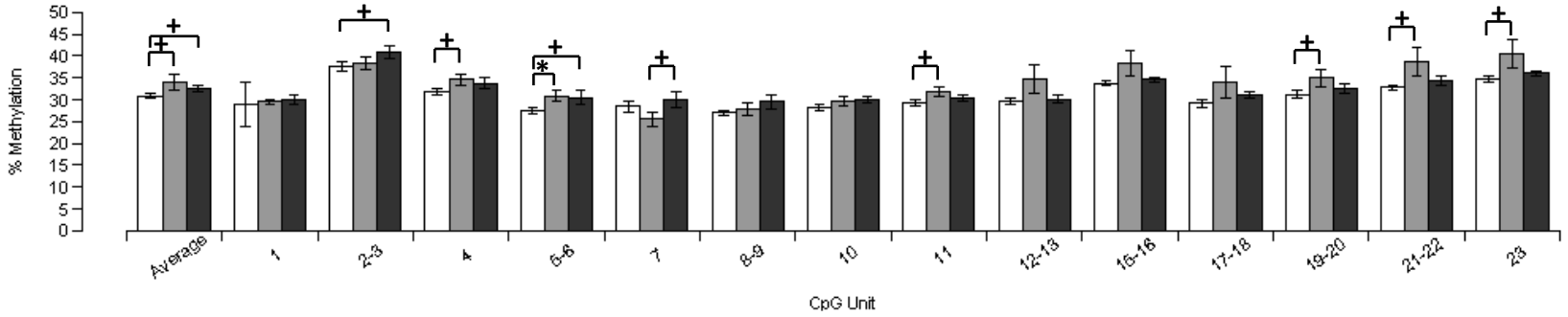


Figure 5.16 Continued

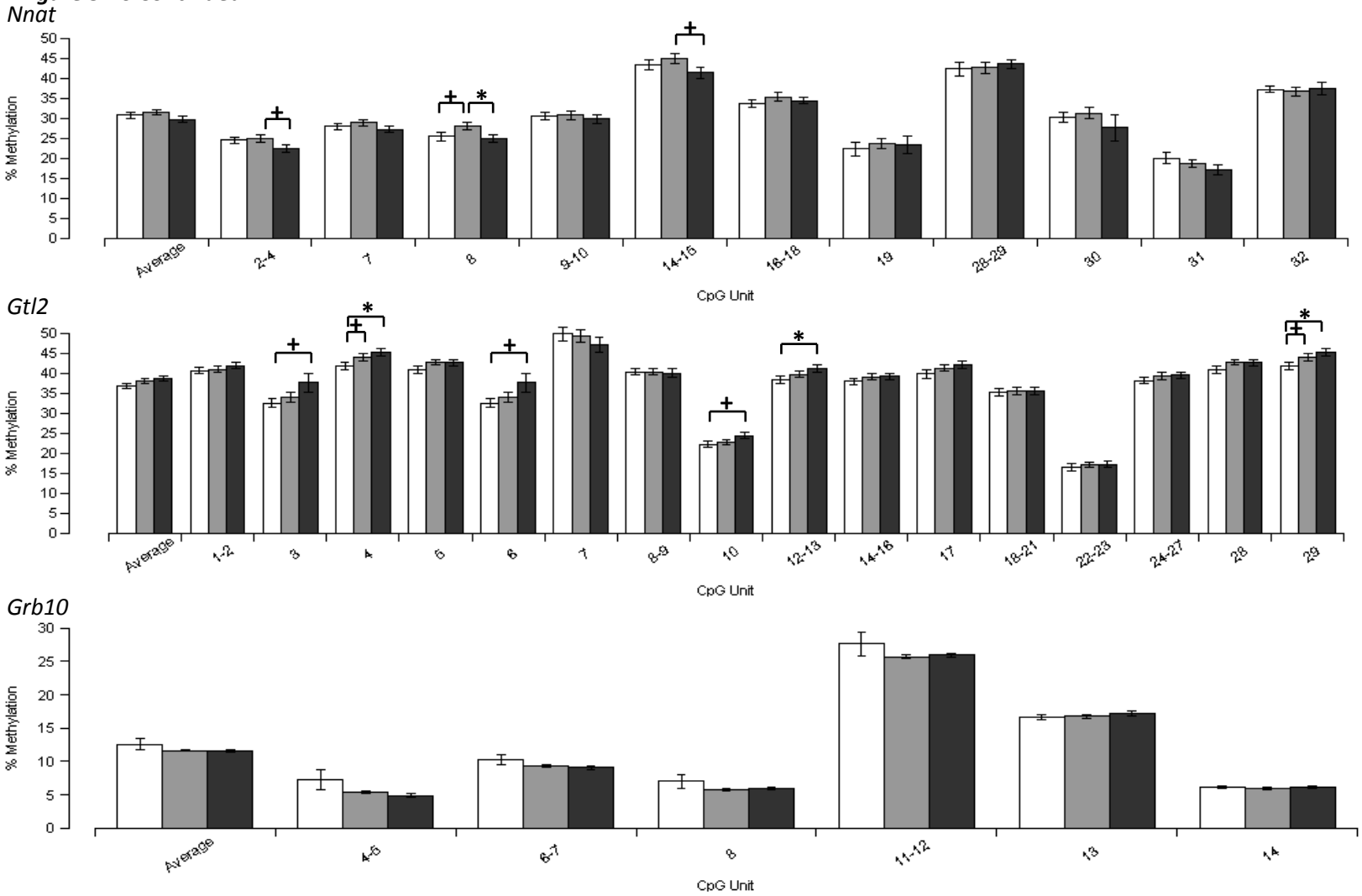
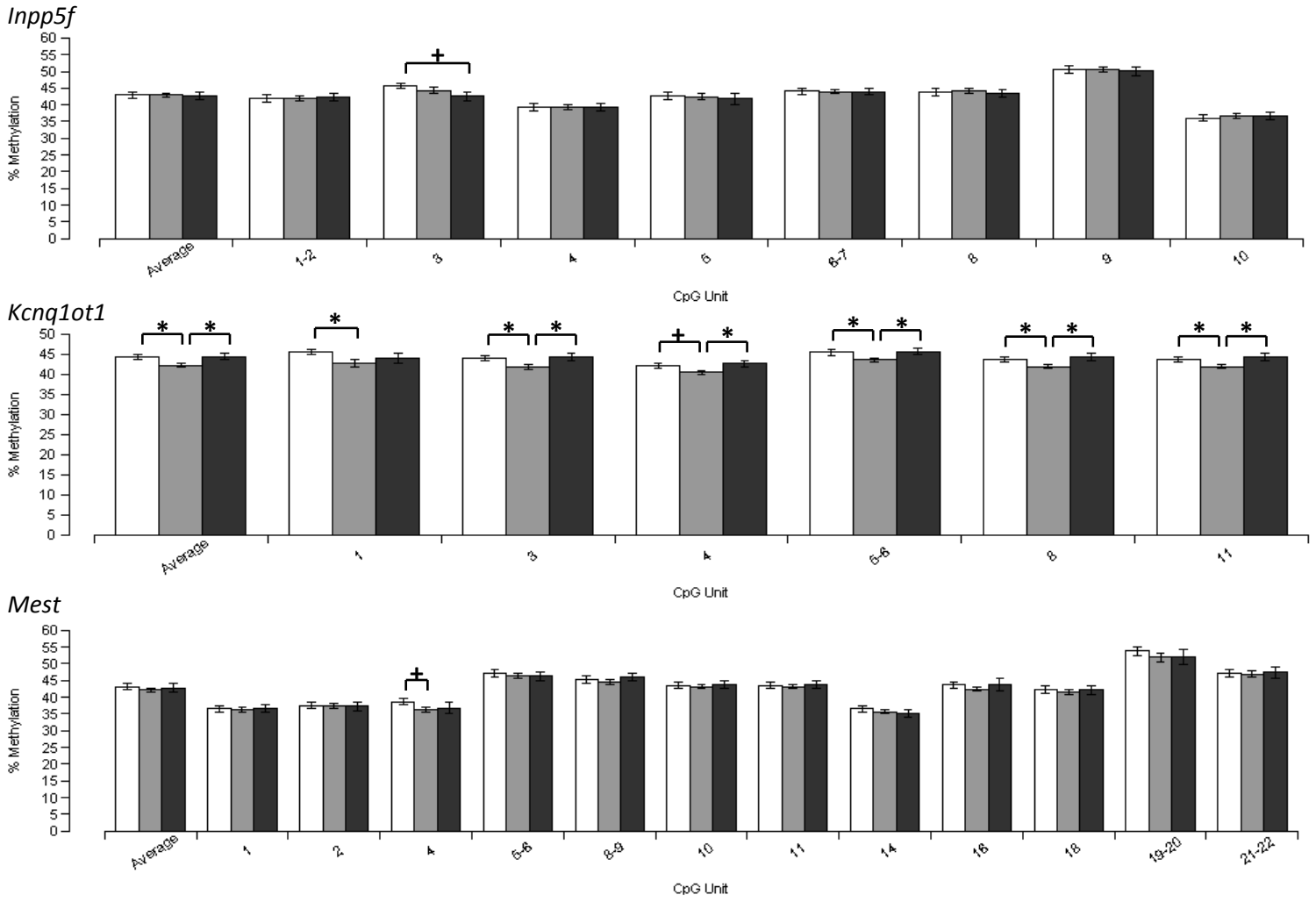


Figure 5.16 Continued



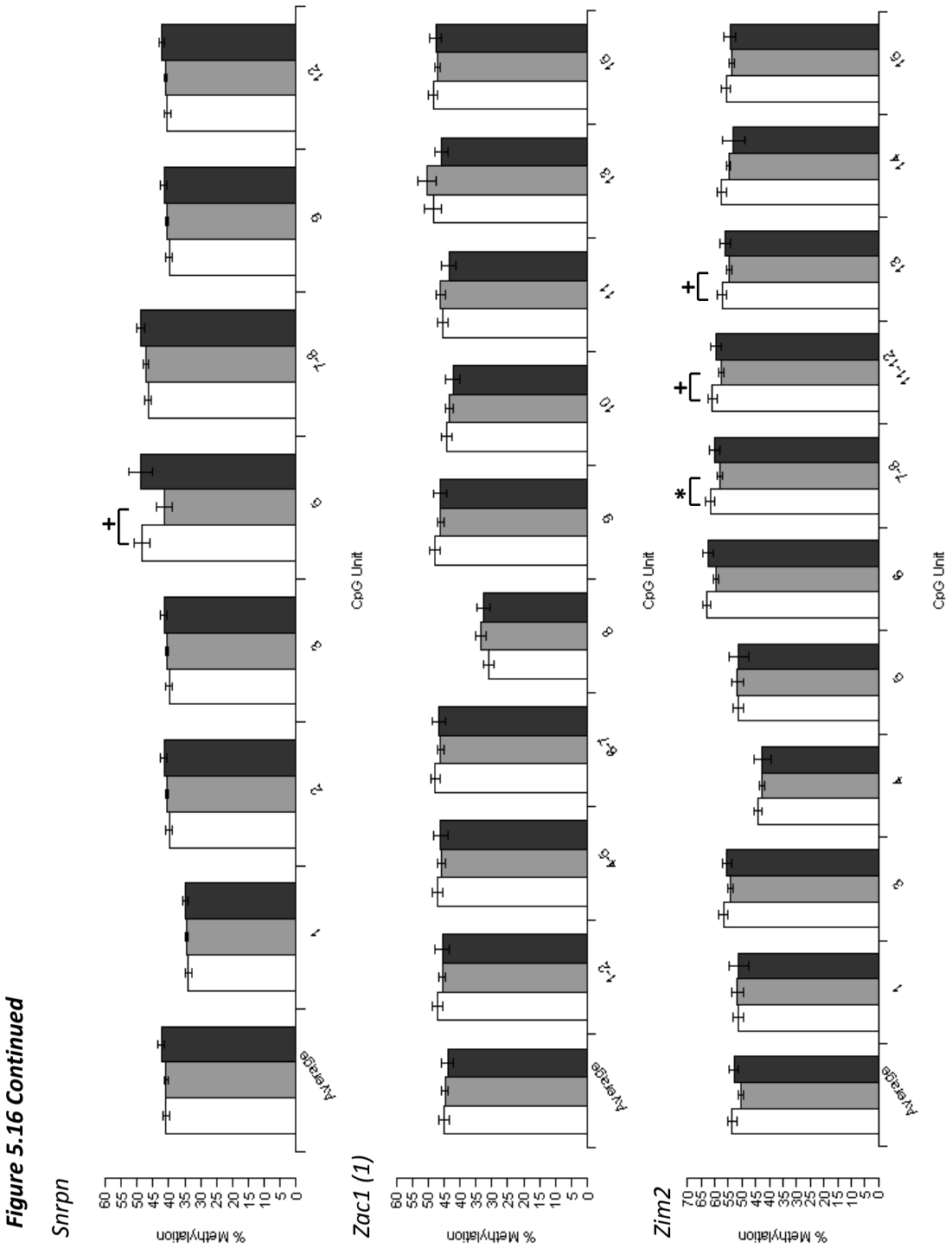


Figure 5.16 Continued

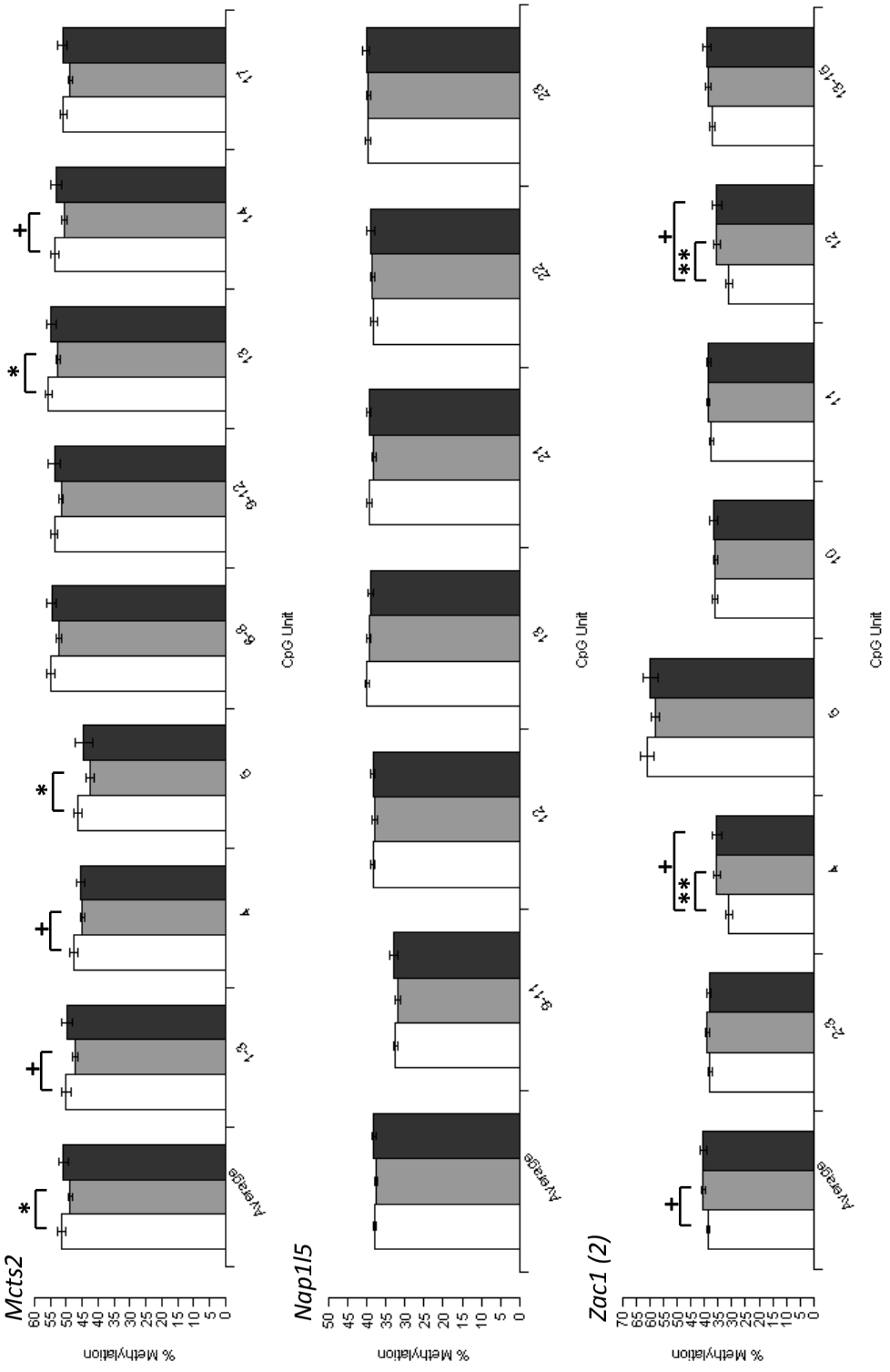


Figure 5.16 Continued

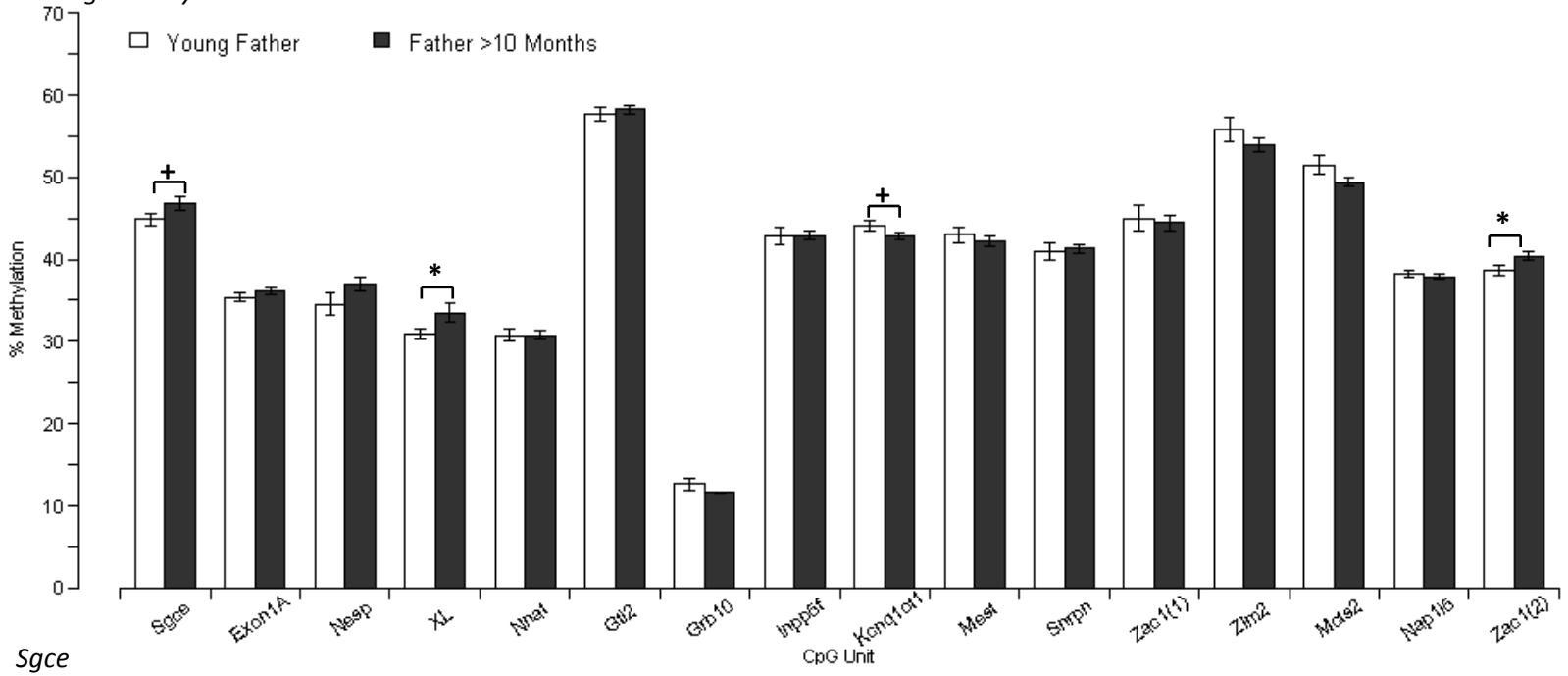
The results for the analyses comparing the offspring of young fathers to the offspring of fathers over 10 months are shown in *Figure 5.17*. There was a significant increase in the average DNA methylation across the PCR amplicon for two DMRs; Gnas XL ($t = -2.03$, d.f. = 50, $p = 0.05$) and Zac1 DMR 2 ($t = -2.10$, d.f. = 50, $p = 0.04$), with a trend towards lower methylation across the Kcnq1ot1 DMR ($t = 1.68$, d.f. = 50, $p = 0.10$). Again many individual CpG sites showed a significant (or a trend towards a significant) difference in DNA methylation between groups and these are summarised in *Table 5.3*. Based on these findings in the cerebellum, Gnas-Nesp, Kcnq1ot1 and Mcts2 DMRs were subsequently assessed in other tissues as they showed the most consistent changes in DNA methylation across the entire or the majority of their DMR between the offspring of young fathers and offspring of old fathers.

Figure 5.17 - Methylation of Imprinted Genes in Cerebellum samples II

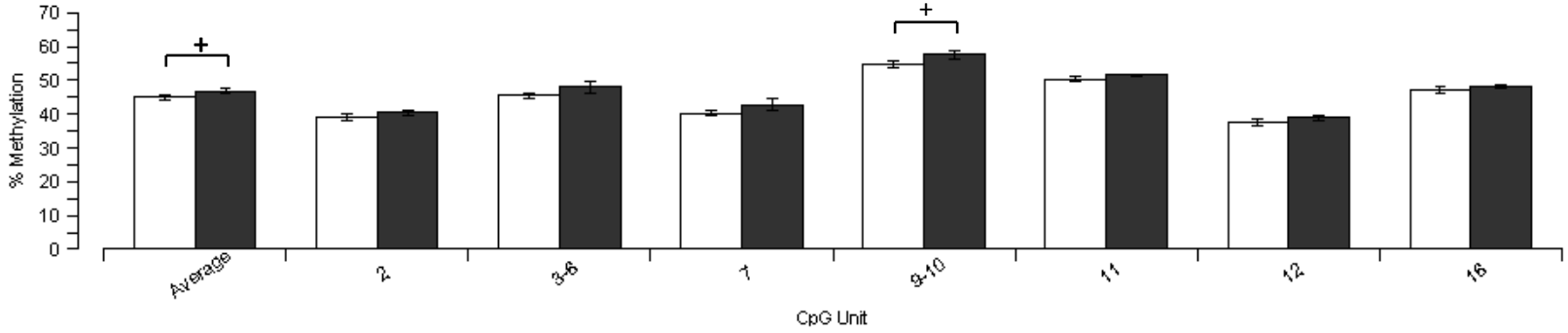
Average methylation in all DMRs and methylation levels across 16 imprinted DMRs in the offspring of young fathers (n = 18) and fathers >10 months (n = 34)

Figure 5.17 Continued

Average Methylation across all DMRs



Sgce



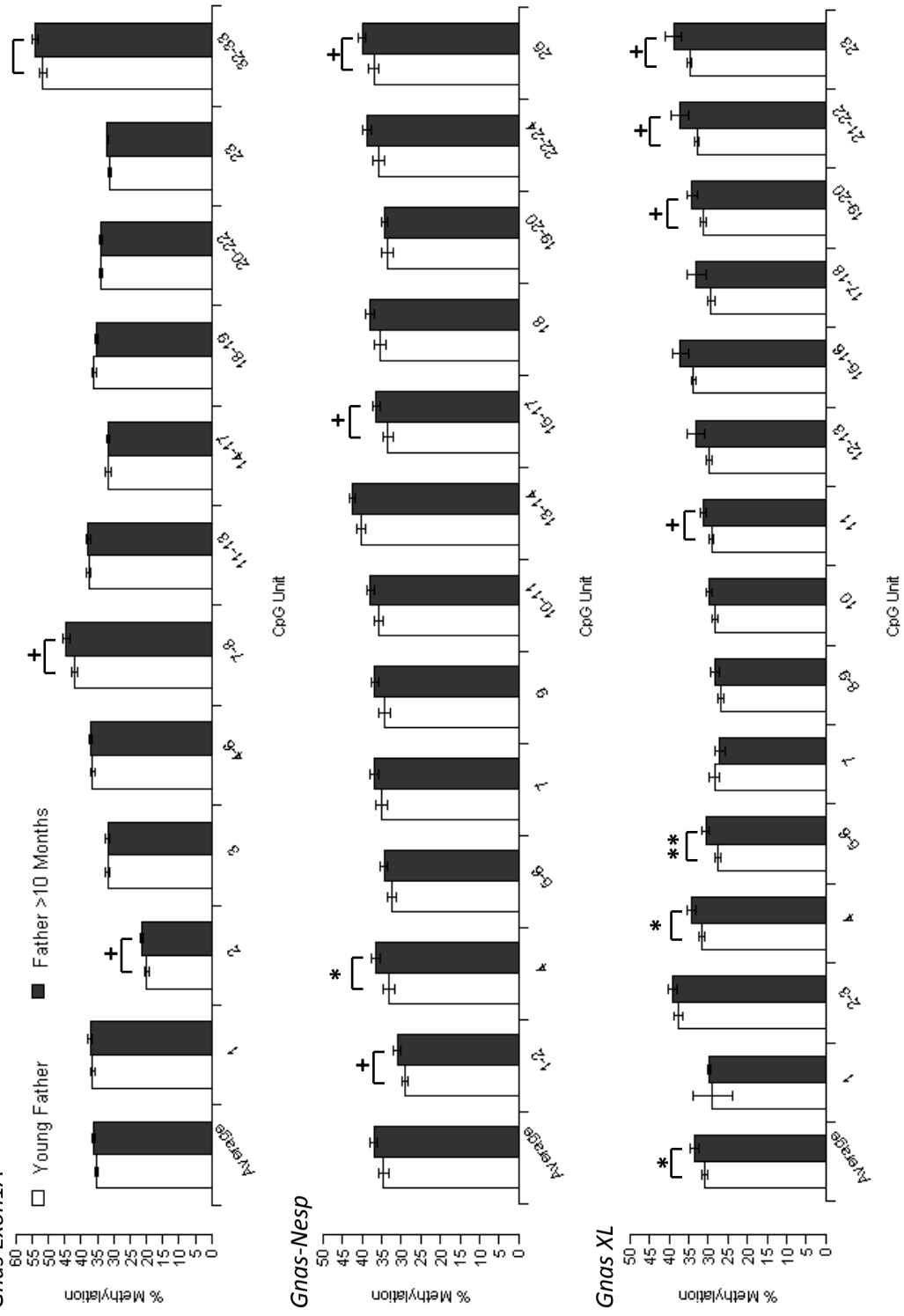


Figure 5.17 Continued

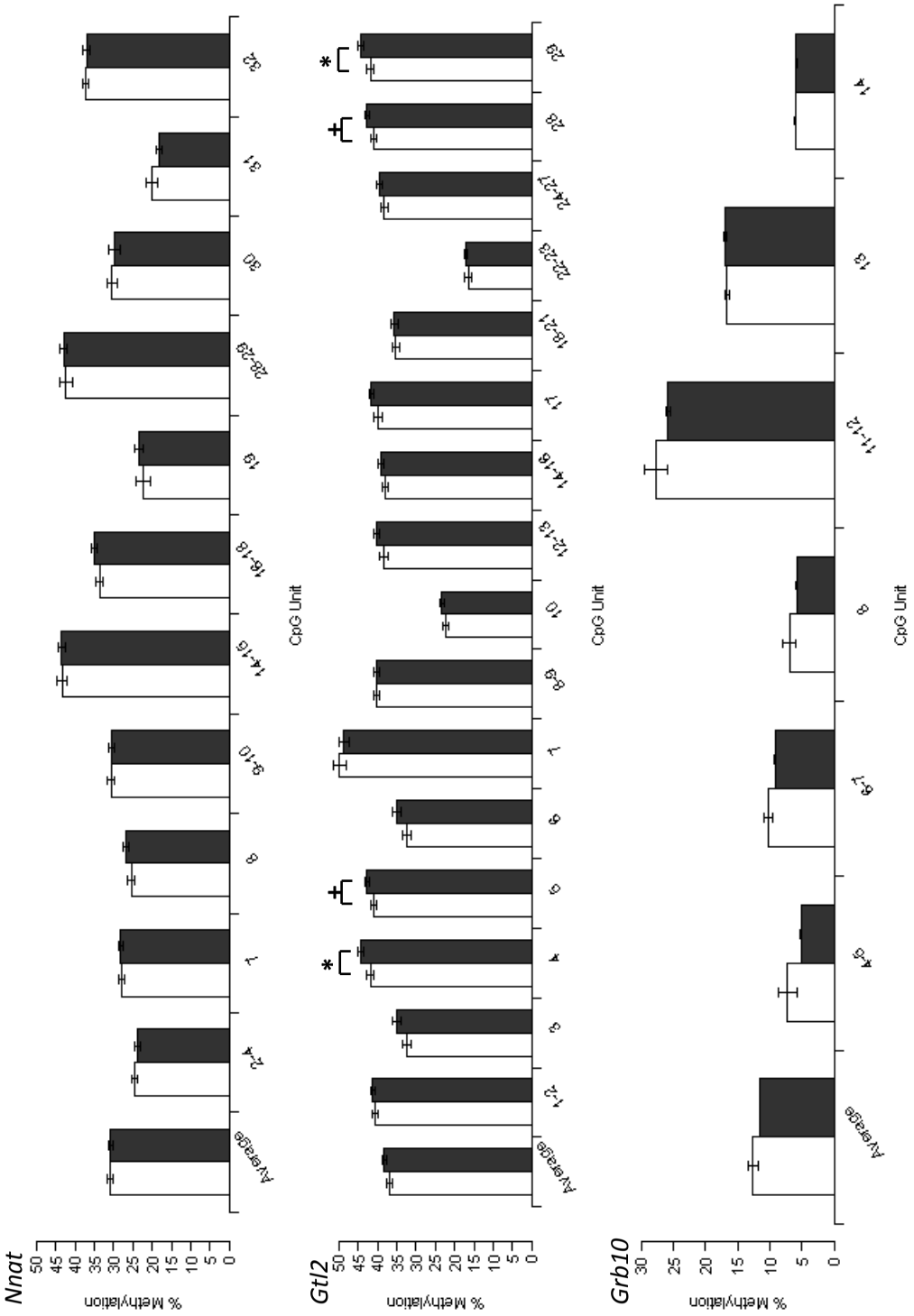


Figure 5.17 Continued

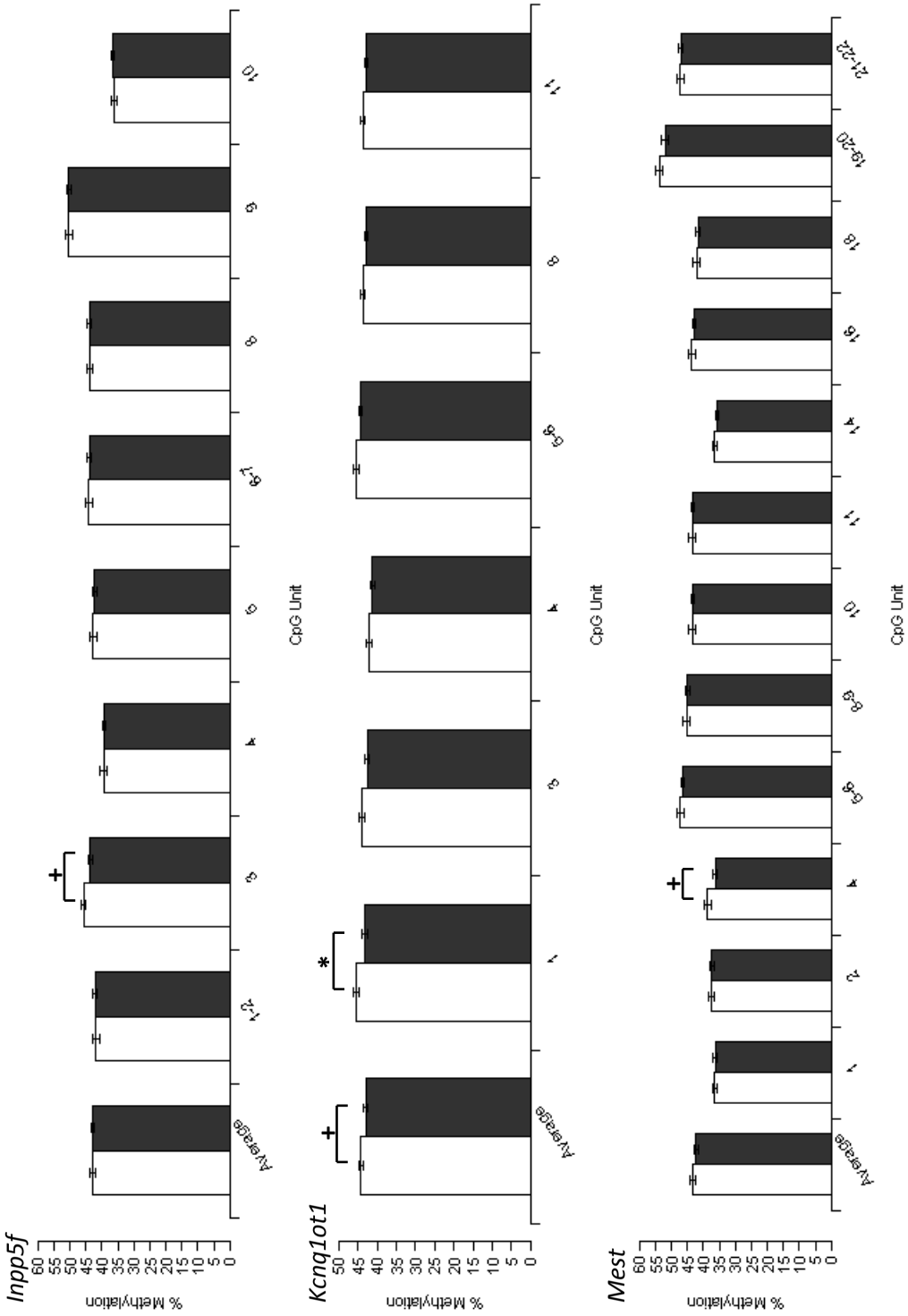


Figure 5.17 Continued

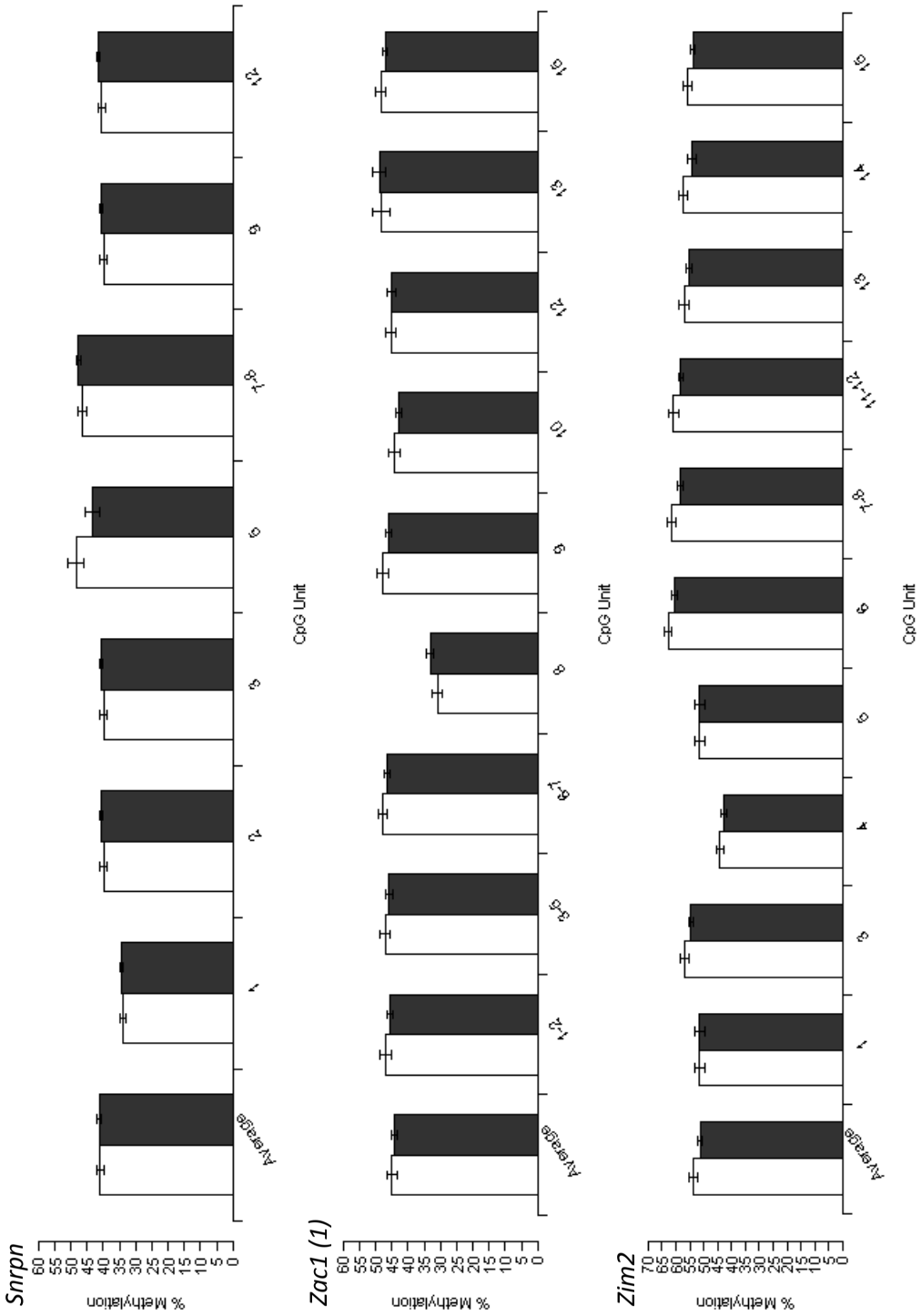
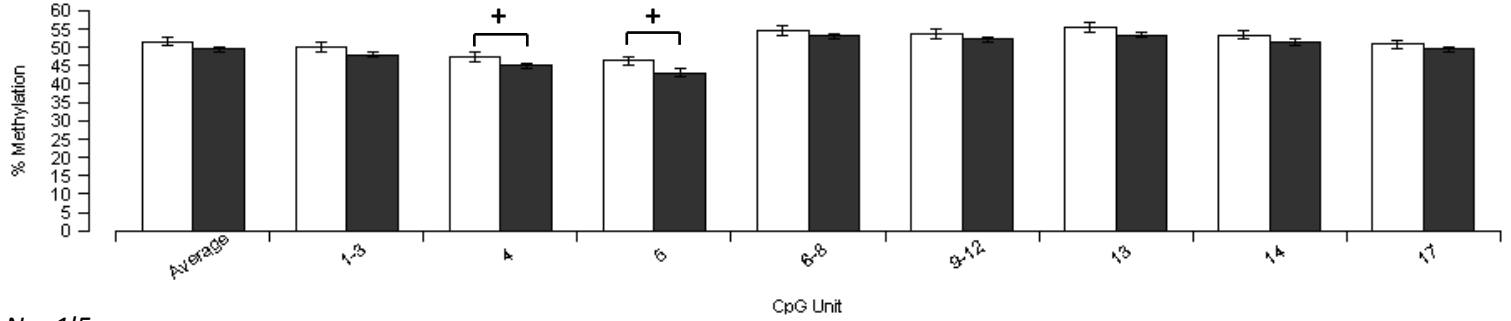


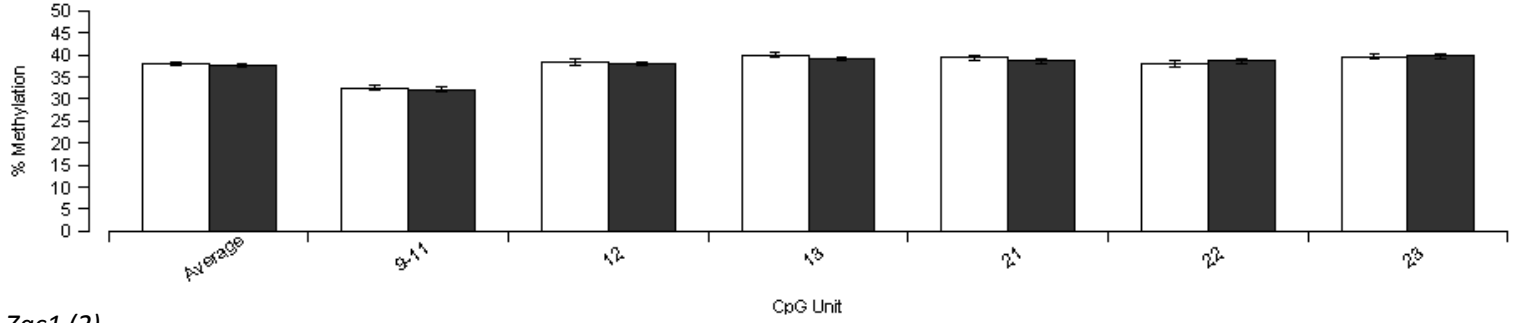
Figure 5.17 Continued

Figure 5.17 Continued

Mcts2



Nap1l5



Zac1 (2)

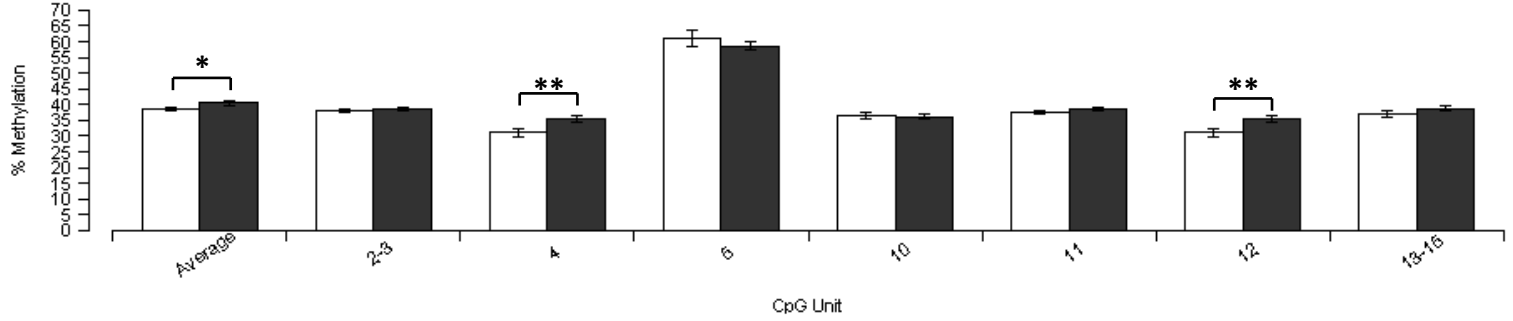


Table 5.3 - Summary of Paternal-Age Associated Differences in DNA Methylation across Imprinting DMRs in the Cerebellum

Numbers refer to CpG unit, p = p-value, t = t-value, 'Direction' refers to direction of change in methylation from first group listed to second group listed (e.g. Offspring of Young (n = 18) vs. Old Father (n = 23) is direction of methylation change for old fathers from methylation in the offspring of young fathers). Filled in boxes indicate where there were no significant or trend towards significance in the results.

DMR	Offspring of Young (n = 18) vs. Old Father (n = 23)					Offspring of Young (n = 18) vs. Very Old Father (n = 11)					Offspring of Old (n = 23) vs. Very Old Father (n = 11)					Offspring of Young vs. >10 Months Father				
	Finding	p	t	% Change	Direction	Finding	p	t	% Change	Direction	Finding	p	t	% Change	Direction	Finding	p	t	% Change	Direction
Sgce	Trend in average	0.09	-1.72	2	↑						Trend in 16	0.06	2.03	2	↓	Trend in average	0.10	-1.67	2	↑
																Trend in 9-10	0.08	-1.79	3	↑
Exon1A	Trend in 2	0.06	-1.99	2	↑	Significant in 7-8	0.05	-2.07	3	↑	Trend in 2	0.1	1.73	1	↓	Trend in 2	0.1	-1.68	1	↑
						Significant in 23	0.01	-2.64	1	↑	Significant in 18-19	0.05	-2.05	2	↑	Trend in 7-8	0.09	-1.75	3	↑
											Trend in 23	0.07	-1.92	1	↑	Trend in 32-33	0.09	-1.76	3	↑
Gnas-Nesp	Trend in average	0.07	-1.84	3	↑						Trend in 1-2	0.07	1.9	3	↓	Trend in 1-2	0.09	-1.73	2	↑
	Significant in 1-2	0.03	-2.33	3	↑						Trend in 13-14	0.06	2.01	3	↓	Significant in 4	0.05	-2.06	3	↑
	Significant in 4	0.03	-2.3	4	↑						Trend in 25	0.09	1.78	3	↓	Trend in 15-17	0.07	-1.89	3	↑
	Trend in 9	0.1	-1.7	3	↑											Trend in 25	0.07	-1.87	3	↑
	Significant in 13-14	0.04	-2.17	4	↑															
	Significant in 15-17	0.03	-2.23	4	↑															
	Trend in 22-24	0.1	-1.67	4	↑															
	Trend in 25	0.1	-1.67	4	↑															

XL	Trend in average	0.1	-1.7	3	↑	Trend in average	0.09	-1.76	2	↑	Trend in 7	0.09	-1.79	4	↑	Significant in average	0.05	-2.03	3	↑
	Trend in 4	0.09	-1.78	3	↑	Trend in 2-3	0.09	-1.8	3	↑						Significant in 4	0.05	-2	2	↑
	Significant in 5-6	0.03	-2.23	3	↑	Trend in 10	0.06	-1.99	2	↑						Highly significant in 5-6	0.01	-2.54	3	↑
	Trend in 11	0.06	-1.93	3	↑											Trend in 11	0.06	-1.98	2	↑
	Trend in 19-20	0.09	-1.79	4	↑											Trend in 19-20	0.07	-1.87	3	↑
	Trend in 21-22	0.1	-1.71	6	↑											Trend in 21-22	0.07	-1.87	4	↑
	Trend in 23	0.1	-1.72	6	↑											Trend in 23	0.07	-1.86	4	↑
Nnat	Trend in 8	0.07	-1.85	3	↑						Trend in 2-4	0.1	1.71	2	↓					
											Significant in 8	0.04	2.14	3	↓					
											Trend in 14-15	0.06	1.96	4	↓					
Gtl2	Trend in 4	0.09	-1.77	2	↑	Trend in 3	0.07	-1.95	5	↑						Significant in 4	0.03	-2.28	3	↑
	Trend in 29	0.09	-1.77	2	↑	Significant in 4	0.02	-2.47	3	↑						Trend in 5	0.08	-1.79	2	↑
						Trend in 6	0.07	-1.95	5	↑						Trend in 28	0.08	-1.79	2	↑
						Trend in 10	0.07	-1.87	2	↑						Significant in 29	0.03	-2.28	3	↑
						Significant in 12-13	0.05	-2.06	3	↑										
						Significant in 29	0.02	-2.47	3	↑										
Grb10																				
Inpp5f						Trend in 3	0.06	2	3	↓						Trend in 3	0.08	1.83	2	↓
Kcnq1ot1	Significant in average	0.02	2.47	2	↓						Significant in average	0.04	-2.22	2	↑	Trend in average	0.1	1.68	1	↓
	Significant in 1	0.02	2.36	3	↓						Significant in 3	0.02	-2.46	3	↑	Significant in 1	0.03	2.29	2	↓
	Significant in 3	0.02	2.44	2	↓						Significant in 4	0.03	-2.3	2	↑					
	Trend in 4	0.06	1.91	2	↓						Significant in 5-6	0.02	-2.46	2	↑					
	Significant in 5-6	0.05	2.04	2	↓						Significant in 8	0.03	-2.49	2	↑					
	Significant in 8	0.03	2.26	2	↓						Significant in 11	0.03	-2.49	2	↑					
	Significant in 11	0.03	2.26	2	↓															
Mest	Trend in 4	0.06	1.96	3	↓											Trend in 4	0.07	1.84	2	↓
Snrpn	Trend in 5	0.06	1.99	7	↓															

Zac1 (1)															
Zim2	Significant in 7-8	0.04	2.14	3	↓										
	Trend in 11-12	0.08	1.81	4	↓										
	Trend in 13	0.1	1.71	3	↓										
Mcts2	Significant in average	0.05	2.09	3	↓							Trend in 4	0.09	1.75	2	↓
	Trend in 1-3	0.09	1.75	3	↓							Trend in 5	0.06	1.91	3	↓
	Trend in 4	0.08	1.8	3	↓											
	Significant in 5	0.03	2.25	4	↓											
	Significant in 13	0.05	2.07	3	↓											
	Trend in 14	0.06	1.92	3	↓											
Nap1l5															
Zac1 (2)	Trend in average	0.06	-1.9	2	↑	Trend in 4	0.07	-1.91	4	↑		Significant in average	0.04	-2.1	2	↑
	Highly significant in 4	0.01	-2.66	4	↑	Trend in 12	0.07	-1.91	4	↑		Highly significant in 4	0.01	-2.77	4	↑
	Highly significant in 12	0.01	-2.66	4	↑							Highly significant in 12	0.01	-2.77	4	↑

Figure 5.18 – Offspring of Young Fathers vs. Offspring of Old Fathers in Cerebellum

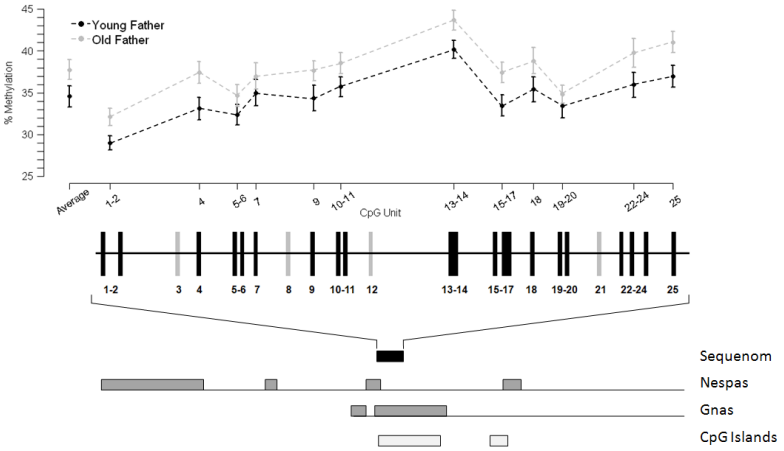
Three DMRs which showed consistent methylation changes across the entire PCR region as well as schematic of the region. Greyed out boxes are CpG units which were not assayed.

a) Gnas-Nesp

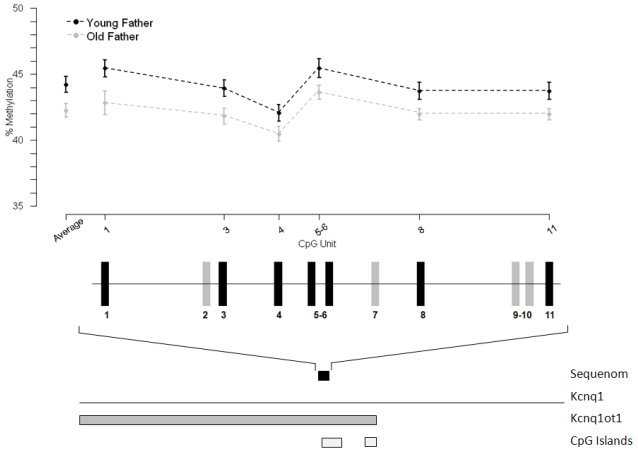
b) Kcnqot1

c) *Mcts2*

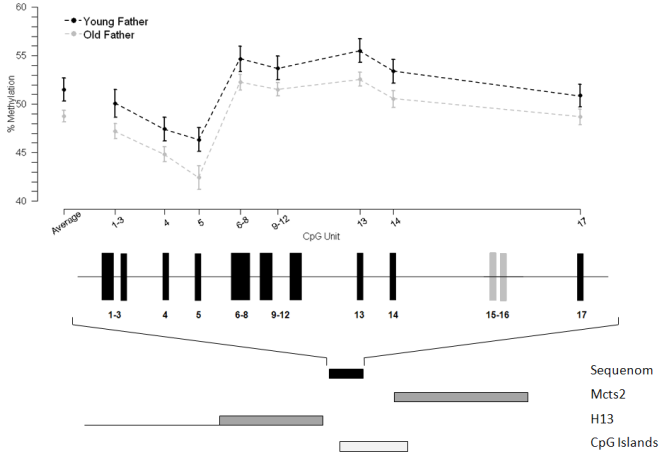
a)



b)

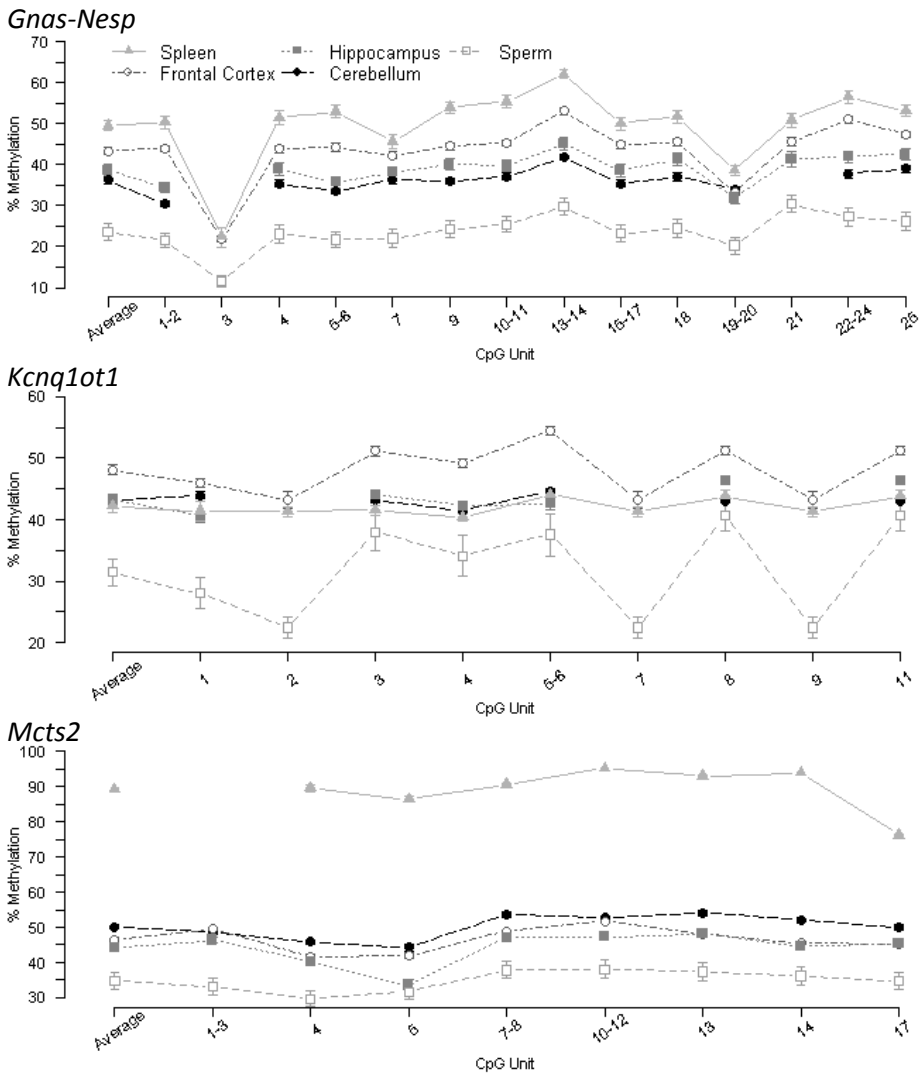


c)



Based on the results from the 16 DMRs assessed in the cerebellum, we next examined the tissue-specificity of differences by examining three specific regions (Mcts2, Gnas-Nesp and Kcnq1ot1) in spleen, frontal cortex, and hippocampal tissue obtained from the same individual animals. Furthermore, we examined these regions in sperm DNA samples obtained from young, old, and very old males. Average levels of DNA methylation across the 3 DMRs in the different tissues are shown in *Figure 5.19*

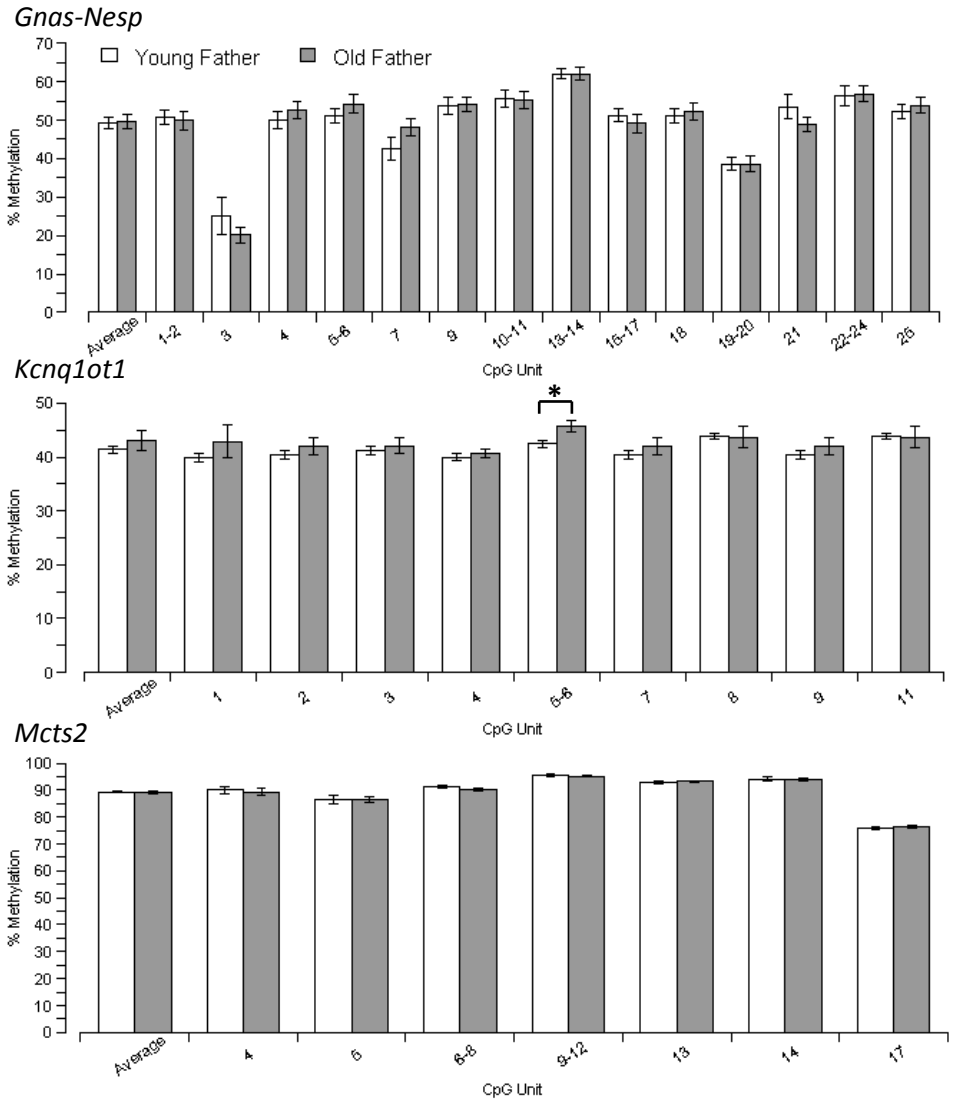
Figure 5.19 – Methylation across Gnas-Nesp, Kcnq1ot1 and Mcts2 in all Tissues



5.5.2.2 Spleen

Compared to other tissues, Mcts2 was observed to be significantly hypermethylated in the spleen (*Figure 5.19*). This concurs with information showing that it is expressed at a very low level in this tissue (EuropeanBioinformaticsInstitute 2011). The results for the Gnas-Nesp and Kcnq1ot1 DMRs are shown in *Figure 5.20*, comparing the offspring of young fathers to the offspring of old fathers. There was no significant difference between overall DNA methylation for either amplicon (Gnas-Nesp ($t = -0.16$, d.f. = 39, $p = 0.87$) and Kcnq1ot1 ($t = -0.81$, d.f. = 39, $p = 0.43$)). In the latter, only one CpG unit (5-6) showed a difference between groups ($t = -2.59$, d.f. = 39, $p = 0.02$) with the offspring of old fathers having a higher level of methylation at this site, which is the opposite direction seen in the cerebellum. No significant differences were observed between individual sites in Mcts2 or in average methylation across the PCR regions ($t = 0.27$, d.f. = 39, p-value = 0.79).

Figure 5.20 - Methylation levels of Imprinted Genes in Spleen

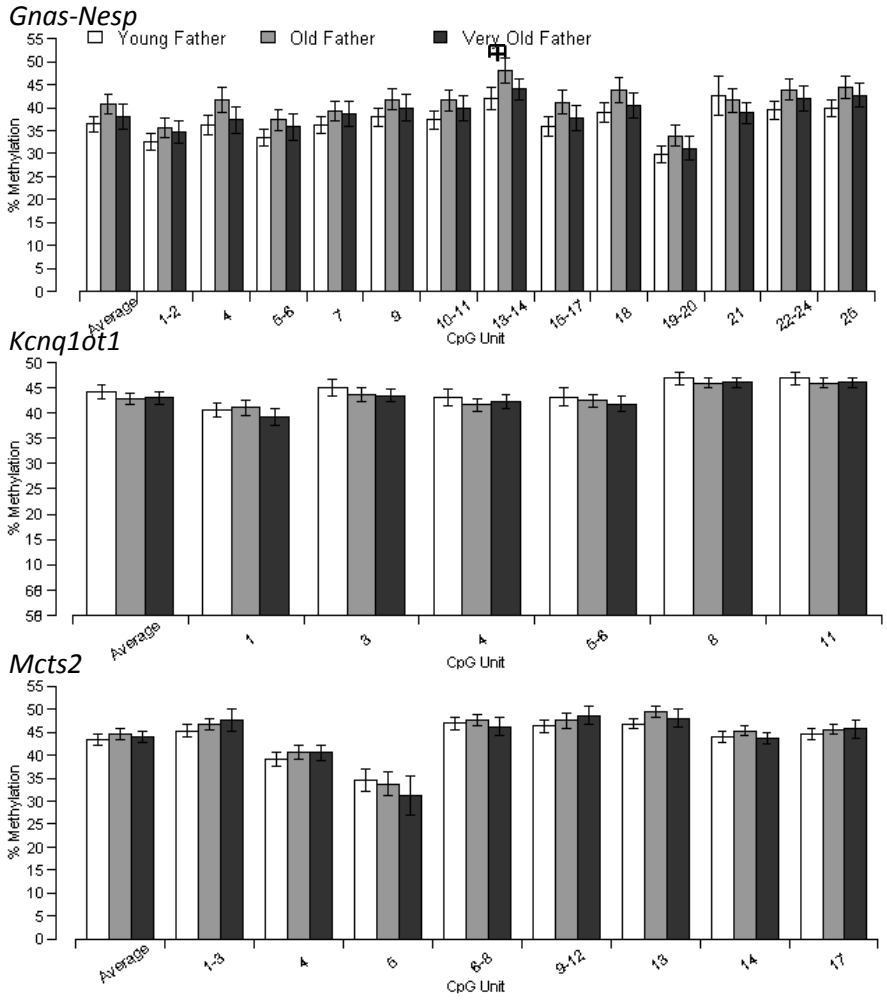


5.5.2.3 Hippocampus

The results of the analysis of the analysis between advancing paternal age and DNA methylation at the Mcts2, Gnas-Nesp and Kcnq1ot1 DMRs in the hippocampus are shown in *Figure 5.21* (for offspring of young fathers, offspring of old fathers and offspring of very old fathers). There were no significant differences between groups for either individual CpG sites or average DNA methylation across any of the regions between all three groups (Kcnq1ot1 ($F(2, 50) = 0.32$, $p = 0.73$), Mcts2 ($F(2, 50) = 0.26$, $p = 0.77$) or Gnas-Nesp ($F(2, 50) = 1.11$, $p = 0.34$)). Only unit 13-14 in Gnas-Nesp showed a trend towards significance between the offspring of young fathers and offspring of old fathers ($t = -1.69$, d.f. = 39 $p = 0.098$). The pattern of group DNA methylation at Gnas-Nesp shows a similar inverted *u* shaped pattern

than in DNA from cerebellum (*Figure 5.16*) although the group differences aren't significant in the hippocampus.

Figure 5.21 - Methylation of Imprinted Genes in Hippocampus Samples

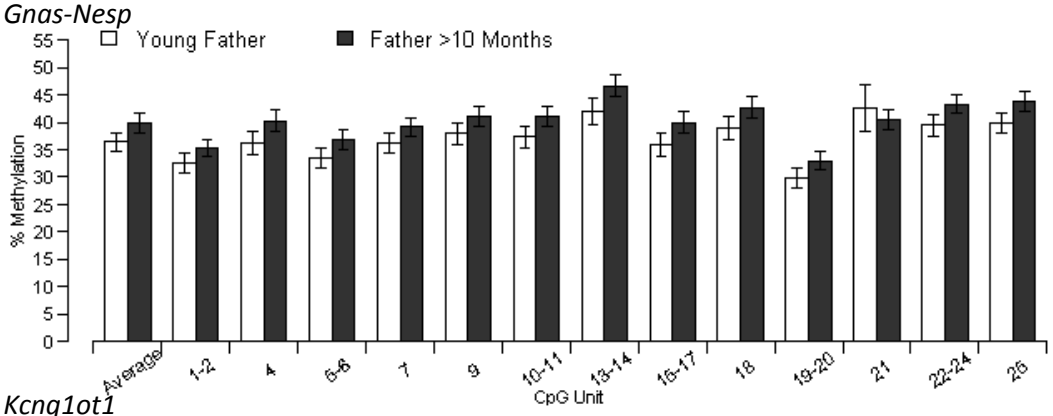


Likewise, there were no significant differences between the offspring of young fathers and offspring of fathers over 10 months for any individual CpG units or amplicon-averaged DNA methylation for any of the DMRs (Kcnq1ot1 ($t = 0.75$, d.f. = 50, $p = 0.46$), Mcts2 ($t = -0.68$, d.f. = 50, $p = 0.50$) or Gnas-Nesp ($t = -1.41$, d.f. = 50, $p = 0.16$)) (*Figure 5.22*).

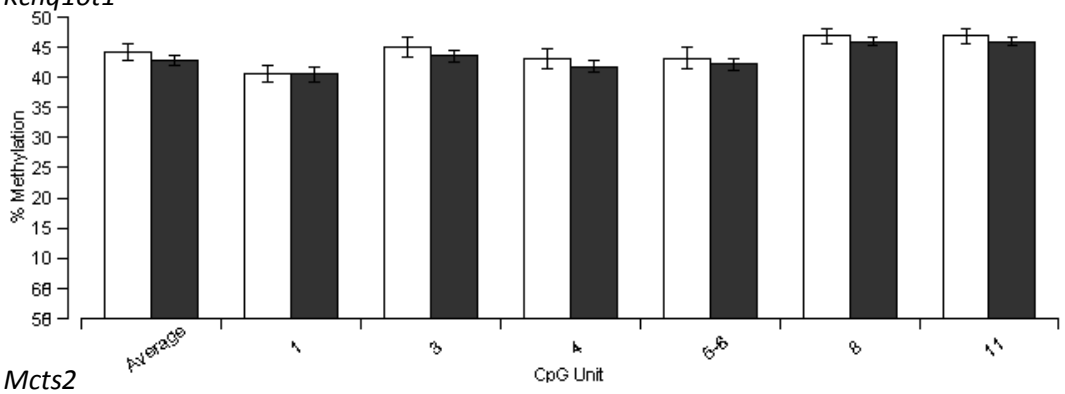
Figure 5.22 – Methylation of Imprinted Genes in Hippocampus Samples II

Offspring of Young Fathers vs. Offspring of Fathers >10 Months

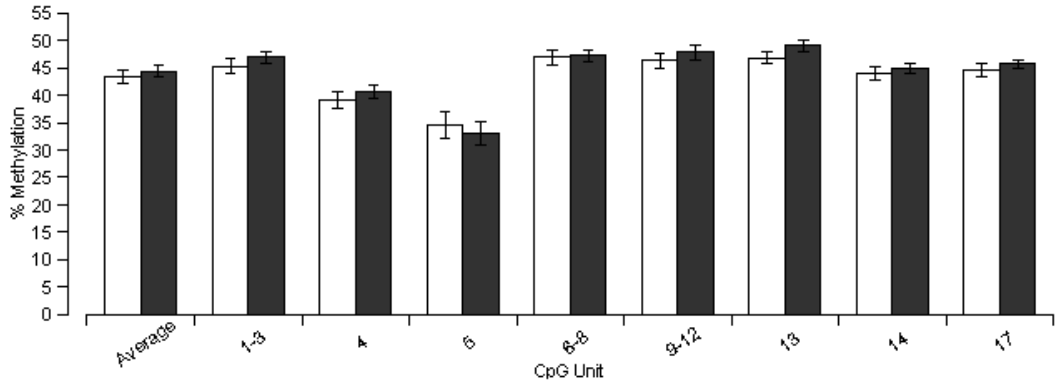
Gnas-Nesp



Kcnq1ot1



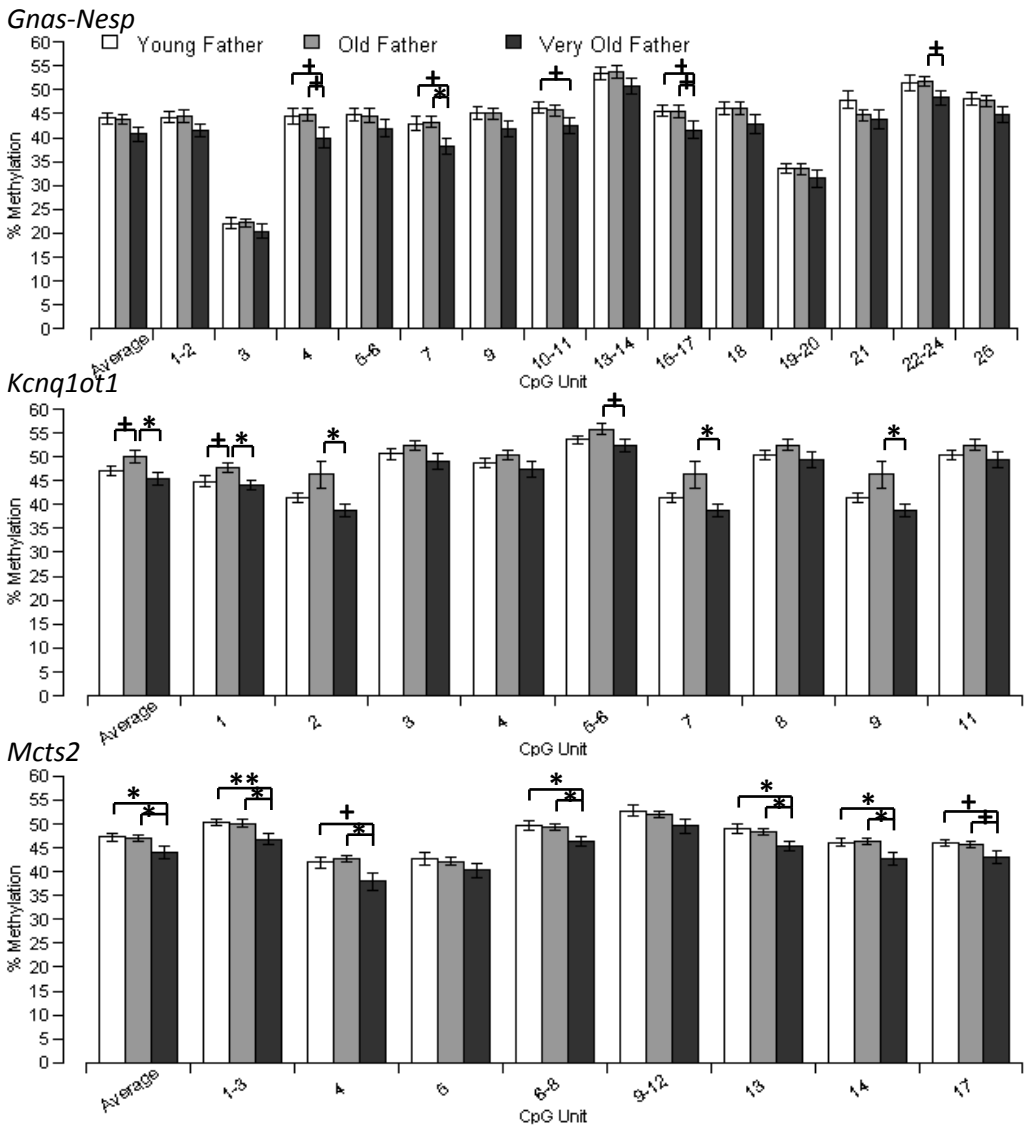
Mcts2



5.5.2.4 Frontal Cortex

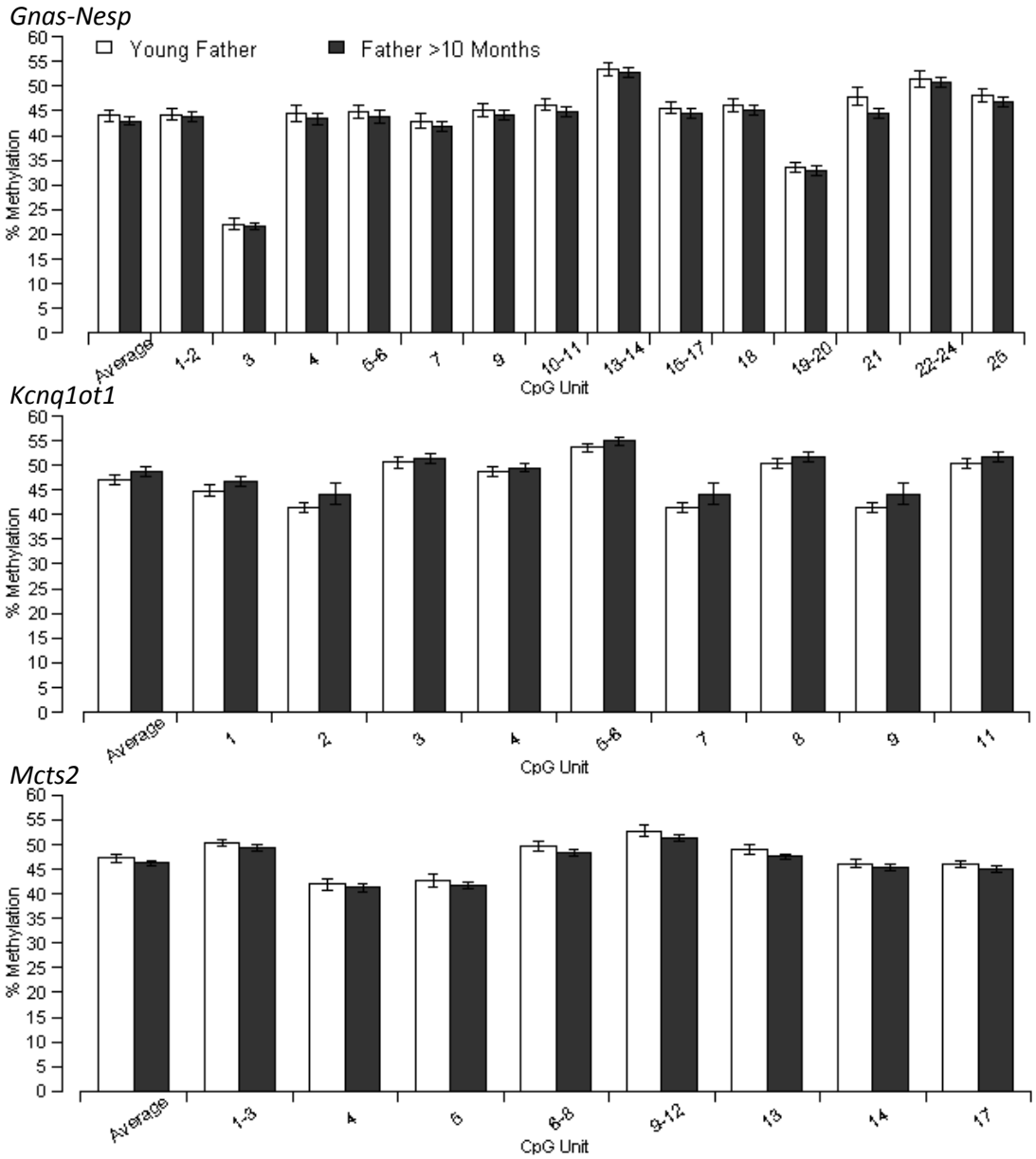
The results of the analysis of the analysis between advancing paternal age and DNA methylation at the Gnas-Nesp, Kcnq1ot1 and Mcts2 DMRs in the frontal cortex are shown in *Figure 5.23* (for the offspring of young fathers, offspring of old fathers and offspring of very old fathers). Interestingly, average DNA methylation was significantly different between groups across the Kcnqot1 DMR (F (2, 50) = 3.43, p = 0.04) and the Mcts2 DMR (F (2, 50) = 3.34, p = 0.04), but there was no significant different across the Gnas-Nesp DMR (F (2, 50) = 1.44, p = 0.25). Of note, the negative correlation between paternal age and DNA methylation observed across the Mcts2 DMR reflects the patterns seen in the cerebellum.

Figure 5.23 - Methylation of Imprinted Genes in Frontal Cortex Samples



No significant differences were seen between the offspring of young fathers and offspring of fathers over 10 months in individual CpG units of average methylation for Gnas-Nesp (t = 0.68, d.f. = 50, p = 0.50), Kcnq1ot1 (t =-1.23, d.f. = 50, p = 0.22) or Mcts2 (t = 1.01, d.f. = 50, p = 0.32).

Figure 5.24 – Methylation of Imprinted Genes in Frontal Cortex II



5.5.2.5 Sperm

Post mortem sperm samples obtained from mice not related to the individuals used in the rest of this study were analysed for DNA methylation levels at the same three DMRs nominated from analyses of the cerebellum. The results of DNA methylation analysis in the sperm from young males (two months), old males (ten months) and very old males (twelve months) are shown in *Figure 5.25*.

Strikingly, there was a highly-significant decrease in average DNA methylation across both the Gnas-Nesp DMR ($F(2, 20) = 17.43$, $p = 4.15 \times 10^{-5}$), and across the Mcts2 DMR ($F(2, 20) = 12.623$, $p = 2.85 \times 10^{-4}$). Although there was no significant difference in average DNA methylation between the groups across the Kcnq1ot1 DMR ($F(2, 20) = 1.50$, $p = 0.25$), three CpG sites were significantly different between young males and old males (sites 2, 7 and 9). These results are further summarised in Table 5.4. Sperm samples from the young fathers were then compared to the combined data from old males and very old males to make the males over ten months old group, paralleling our analysis on tissue from the test animals. Again, there was a significant decrease in DNA methylation across the Gnas-Nesp DMR ($t = 4.39$, $d.f. = 21$, $p = 1.64 \times 10^{-3}$) and the Mcts2 DMR ($t = 3.97$, $d.f. = 21$, $p = 3.93 \times 10^{-3}$). There was no significant difference in the level of average DNA methylation across the Kcnq1ot1 DMR ($t = 1.66$, $d.f. = 21$, $p = 0.13$). This data in shown in *Figure 5.26* and summarised in *Table 5.5.*

Figure 5.25 – Methylation of Imprinted Gene DMRs in Sperm Samples Obtained from Young, Old and Very Old Mice

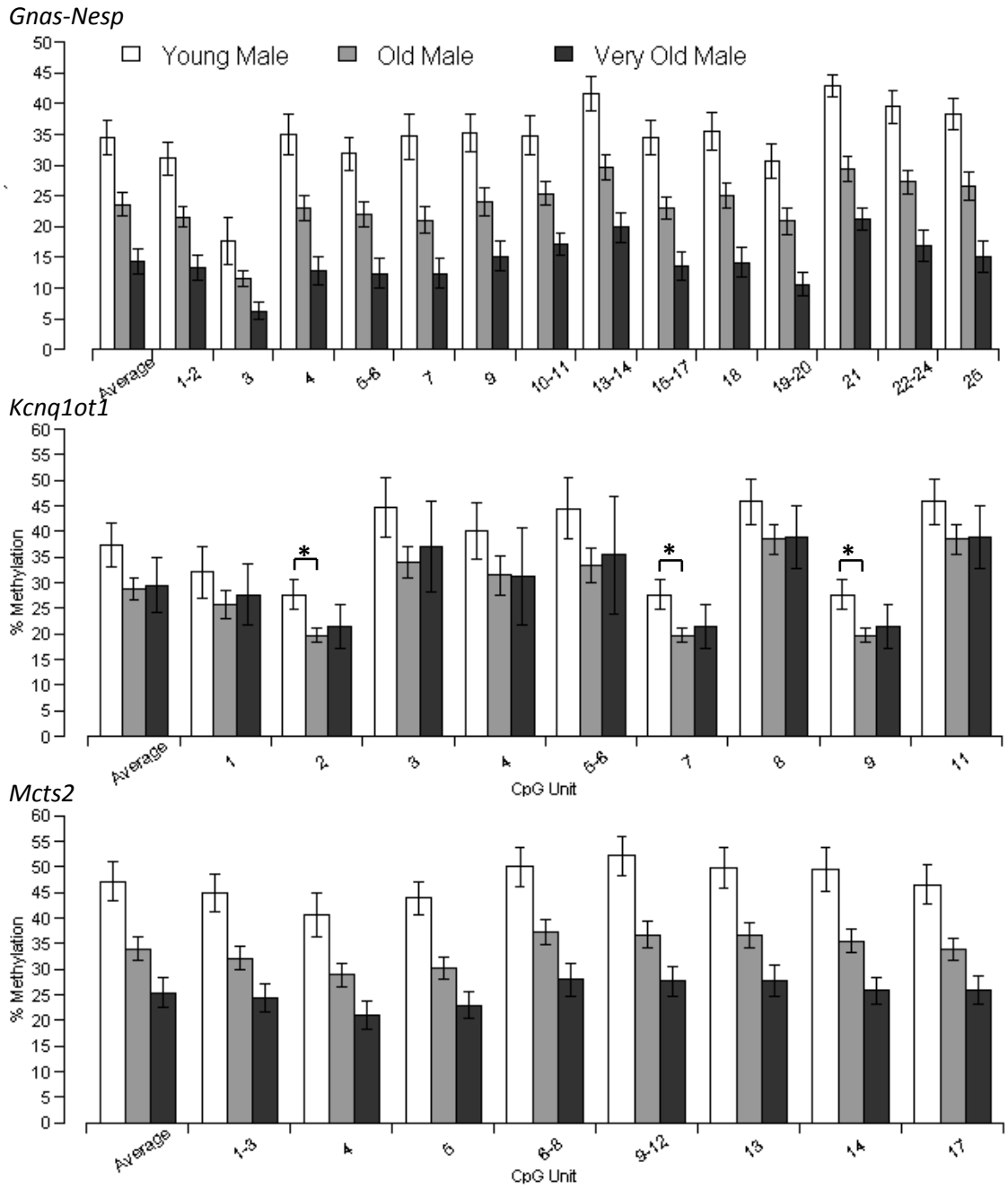


Table 5.4 – Results of Methylation Analysis on Sperm Samples

Numbers refer to CpG unit, p = p-value, t = t-value, 'Direction' refers to direction of change in methylation from first group listed to second group listed (e.g. Young (n = 18) vs. Old Male (n = 23) is direction of change in methylation for old males from methylation in young males)

DMR	Young (n = 6) vs. Old Male (n = 10)					Young (n = 6) vs. Very Old Male (n = 7)					Old (n = 10) vs. Very Old Male (n = 7)				
	Finding	p	t	% Change	Direction	Finding	p	t	% Change	Direction	Finding	p	t	% Change	Direction
Gnas-Nesp	Significant in Average	0.01	3.2	11	↓	Significant in Average	2.35E-04	5.65	20	↓	Significant in Average	0.01	3.24	9	↓
	Significant in 1-2	0.01	3.04	10	↓	Significant in 1-2	3.74E-04	5.29	18	↓	Significant in 1-2	0.01	3.14	8	↓
	Significant in 4	0.01	3.12	12	↓	Significant in 3	0.03	2.8	11	↓	Significant in 3	0.02	2.67	5	↓
	Significant in 5-6	0.02	2.89	10	↓	Significant in 4	2.86E-04	5.63	22	↓	Significant in 4	0.01	3.27	10	↓
	Significant in 7	0.01	3.2	14	↓	Significant in 5-6	2.89E-04	5.33	19	↓	Significant in 5-6	0.01	3.07	10	↓
	Significant in 9	0.01	2.94	11	↓	Significant in 7	7.55E-04	4.97	22	↓	Significant in 7	0.02	2.66	9	↓
	Significant in 10-11	0.03	2.58	9	↓	Significant in 9	4.75E-04	5.11	20	↓	Significant in 9	0.02	2.71	9	↓
	Significant in 13-14	0.01	3.53	12	↓	Significant in 10-11	1.07E-03	4.88	18	↓	Significant in 10-11	0.01	3.07	8	↓
	Significant in 15-17	0.01	3.34	11	↓	Significant in 13-14	1.29E-04	5.84	22	↓	Significant in 13-14	0.01	3.03	10	↓
	Significant in 18	0.02	2.9	10	↓	Significant in 15-17	2.16E-04	5.65	21	↓	Significant in 15-17	0.01	3.28	10	↓
	Significant in 19-20	0.02	2.73	10	↓	Significant in 18	2.46E-04	5.5	21	↓	Significant in 18	4.34E-03	3.47	11	↓
	Significant in 21	1.99E-04	5	13	↓	Significant in 19-20	2.14E-04	5.97	20	↓	Significant in 19-20	2.82E-03	3.57	10	↓
	Significant in 22-24	4.06E-03	3.73	12	↓	Significant in 21	3.61E-06	8.56	22	↓	Significant in 21	0.01	3.04	8	↓
	Significant in 25	0.01	3.43	12	↓	Significant in 22-24	8.34E-05	6.1	23	↓	Significant in 22-24	0.01	3.28	10	↓
						Significant in 25	6.39E-05	6.27	23	↓	Significant in 25	4.98E-03	3.37	11	↓
Kcnq1ot1	Significant in 2	0.04	2.51	8	↓										
	Significant in 7	0.04	2.51	8	↓										
	Significant in 9	0.04	2.51	8	↓										
Mcts2	Significant in Average	0.01	3.04	13	↓	Significant in Average	9.41E-04	4.68	22	↓	Significant in Average	0.04	2.33	8	↓
	Significant in 1-3	0.02	2.97	13	↓	Significant in 1-3	1.27E-03	4.49	21	↓	Significant in 1-3	0.05	2.19	8	↓
	Significant in 4	0.04	2.4	12	↓	Significant in 4	4.58E-03	3.79	20	↓	Significant in 4	0.05	2.17	8	↓
	CpG_5	0.01	3.61	14	↓	Significant in 5	4.54E-04	5.12	21	↓	Significant in 5	0.05	2.2	7	↓
	Significant in 6-8	0.02	2.82	13	↓	Significant in 6-8	1.11E-03	4.47	22	↓	Significant in 6-8	0.04	2.29	9	↓

	CpG_9.10.11.12	0.01	3.39	15	↓	Significant in 9-12	4.87E-04	5.1	24	↓	Significant in 9-12	0.03	2.36	9	↓
	Significant in 13	0.02	2.88	13	↓	Significant in 13	1.15E-03	4.49	22	↓	Significant in 13	0.04	2.23	9	↓
	Significant in 14	0.02	2.91	14	↓	Significant in 14	1.14E-03	4.76	24	↓	Significant in 14	0.02	2.72	10	↓
	Significant in 17	0.02	2.81	13	↓	Significant in 17	1.71E-03	4.36	21	↓	Significant in 17	0.04	2.24	8	↓

Figure 5.26 - Methylation of Imprinted Gene DMRs in Sperm Samples Obtained from Young and >10 Month Old Mice

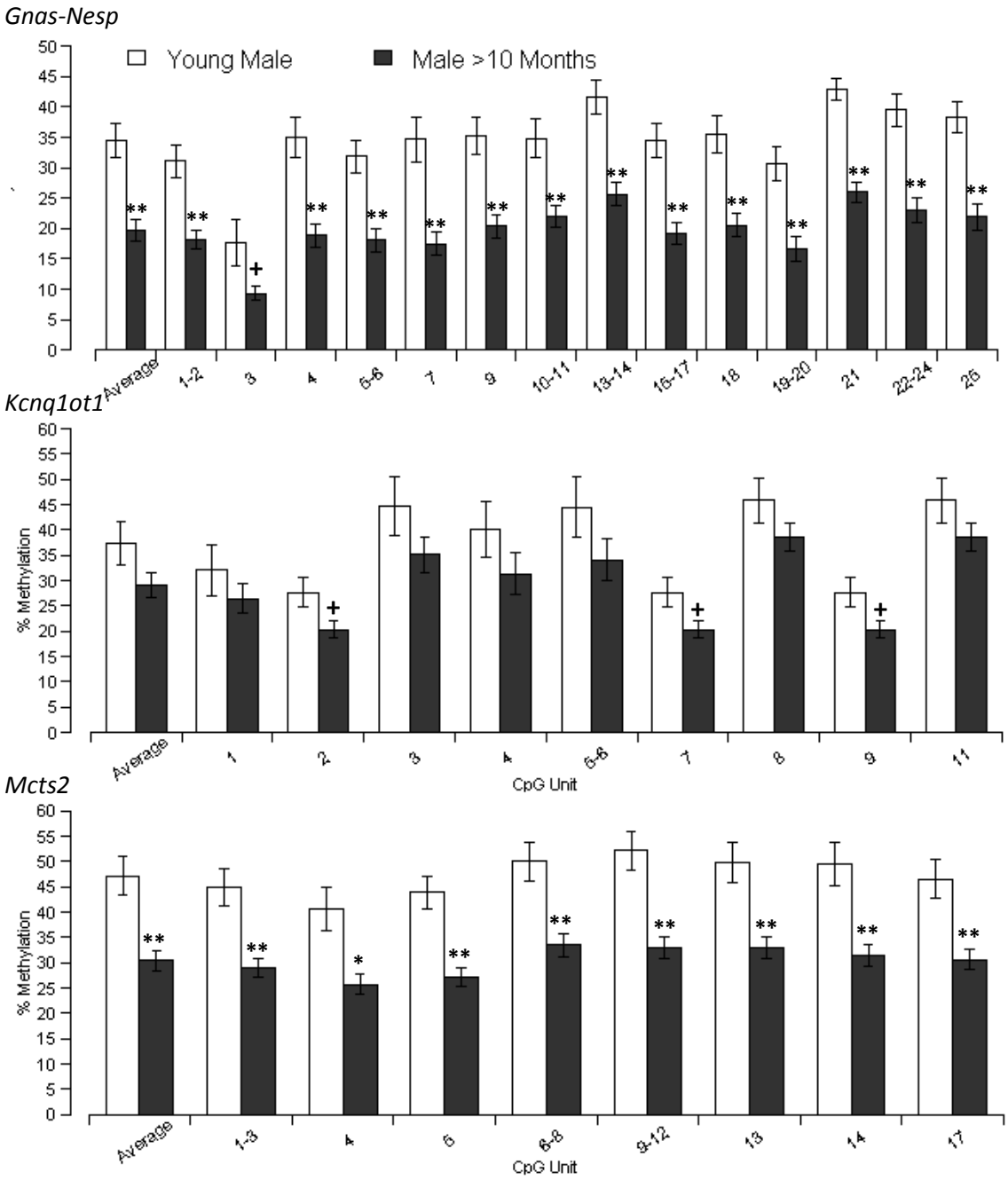


Table 5.5 - Results of Methylation Analysis on Sperm Samples

Numbers refer to CpG unit, p = p-value, t = t-value, 'Direction' refers to direction of change in methylation from first group listed to second group listed

DMR	Offspring of Young (n = 6) vs. >10 Months Father (n = 17)				
	Finding	**p**	**t**	**% Change**	**Direction**
Gnas-Nesp	Significant in Average	1.64E-03	4.39	15	↓
	Significant in 1-2	2.50E-03	4.16	13	↓
	Trend in 3	0.08	2.1	8	↓
	Significant in 4	1.99E-03	4.3	16	↓
	Significant in 5-6	1.99E-03	4.12	14	↓
	Significant in 7	3.55E-03	4.13	17	↓
	Significant in 9	2.69E-03	4.07	15	↓
	Significant in 10-11	0.01	3.64	13	↓
	Significant in 13-14	6.99E-04	4.76	16	↓
	Significant in 15-17	1.47E-03	4.48	15	↓
	Significant in 18	2.09E-03	4.13	15	↓
	Significant in 19-20	2.00E-03	4.11	14	↓
	Significant in 21	6.92E-06	6.88	17	↓
	Significant in 22-24	4.45E-04	4.97	17	↓
	Significant in 25	3.82E-04	4.85	16	↓
Kcnq1ot1	Trend in 2	0.06	2.14	7	↓
	Trend in 7	0.06	2.14	7	↓
	Trend in 9	0.06	2.14	7	↓
Mcts2	Significant in Average	3.93E-03	3.97	17	↓
	Significant in 1-3	4.86E-03	3.85	16	↓
	Significant in 4	0.02	3.14	15	↓
	Significant in 5	1.68E-03	4.56	17	↓
	Significant in 6-8	4.46E-03	3.79	17	↓
	Significant in 9-12	2.04E-03	4.38	19	↓
	Significant in 13	4.55E-03	3.82	17	↓
	Significant in 14	0.01	3.84	18	↓
	Significant in 17	0.01	3.67	16	↓

5.5.3 Transposable Elements

DNA methylation across IAP and LINE-1 repetitive elements was assessed in DNA from cerebellum and frontal cortex from the offspring of young, old and very old fathers and in sperm samples from unrelated males of different ages. Results of average DNA methylation across the assay by tissue and group are summarised in *Table 5.6*.

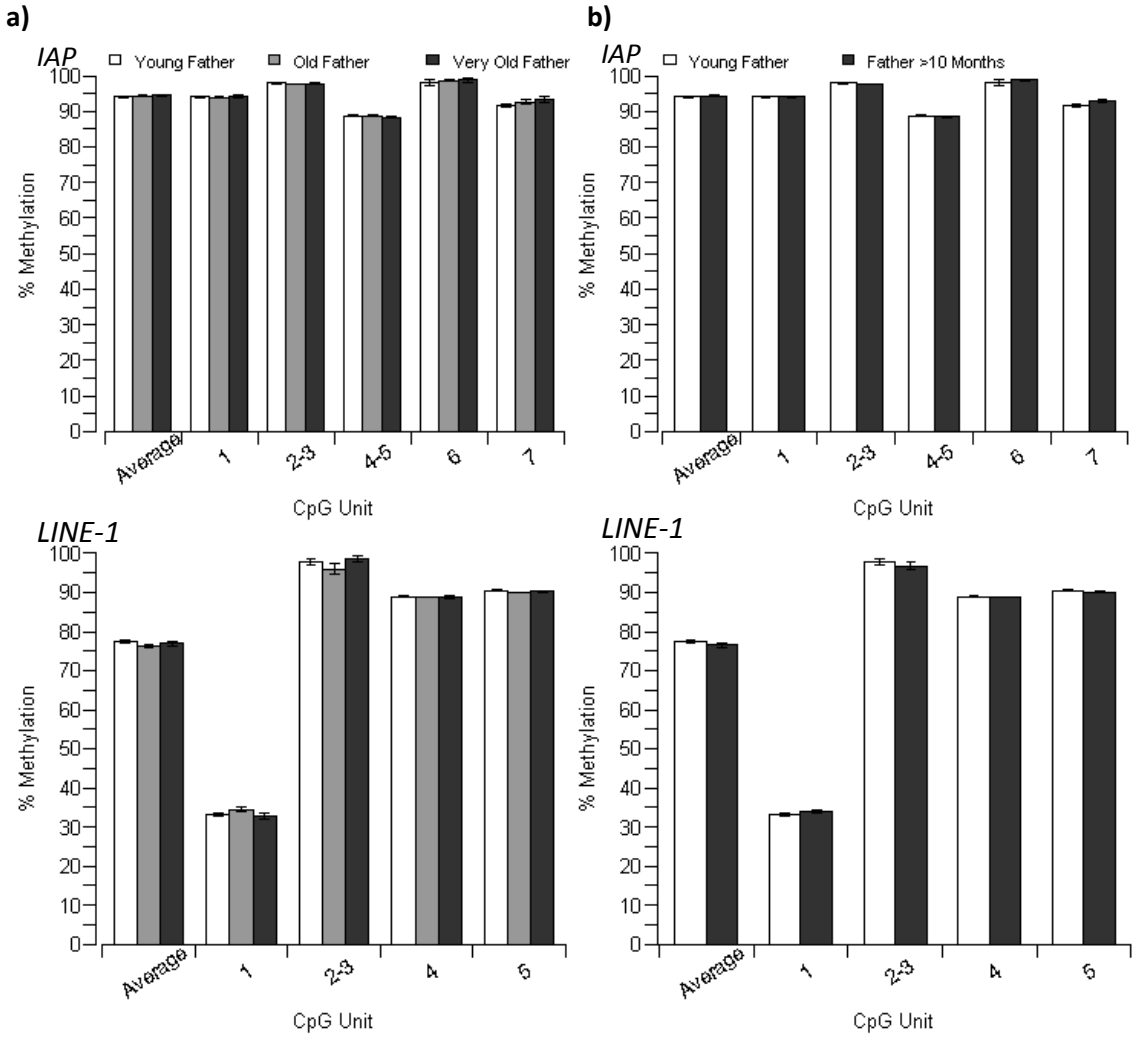
Table 5.6 – Results of Average Methylation from the IAP and LINE-1 Assay

Results of the IAP and LINE-1 Sequenom EpiTYPER assay in DNA from cerebellum, frontal cortex and sperm. Average methylation across the assay in all samples and average methylation across the assay by group are displayed. P-values and t-values between groups are also displayed.

Assay	Tissue	Average	Average Young	Average Old	Average Very Old	Average >10Month	Young v Old		Young v Very Old		Old v Very Old		Young v >10 months	
							p-value	t	p-value	t	p-value	t	p-value	t
IAP	Cerebellum	94.3	94.1	94.4	94.5	94.4	0.40	-0.85	0.29	-1.08	0.71	-0.38	0.31	-1.04
	Frontal Cortex	94.4	94.5	94.1	94.8	94.3	0.58	0.55	0.60	-0.53	0.27	-1.12	0.75	0.32
	Sperm	93.7	93.2	93.8	93.8	93.8	0.46	-0.78	0.46	-0.76	1.00	-0.01	0.42	-0.85
LINE-1	Cerebellum	76.8	77.3	76.3	76.9	76.5	0.13	1.56	0.64	0.48	0.45	-0.78	0.16	1.44
	Frontal Cortex	75.6	75.9	75.4	75.8	75.5	0.23	1.23	0.94	0.07	0.24	-1.20	0.28	1.09
	Sperm	79.0	78.8	78.9	79.2	79.1	0.72	-0.37	0.33	-1.06	0.35	-0.97	0.53	-0.67

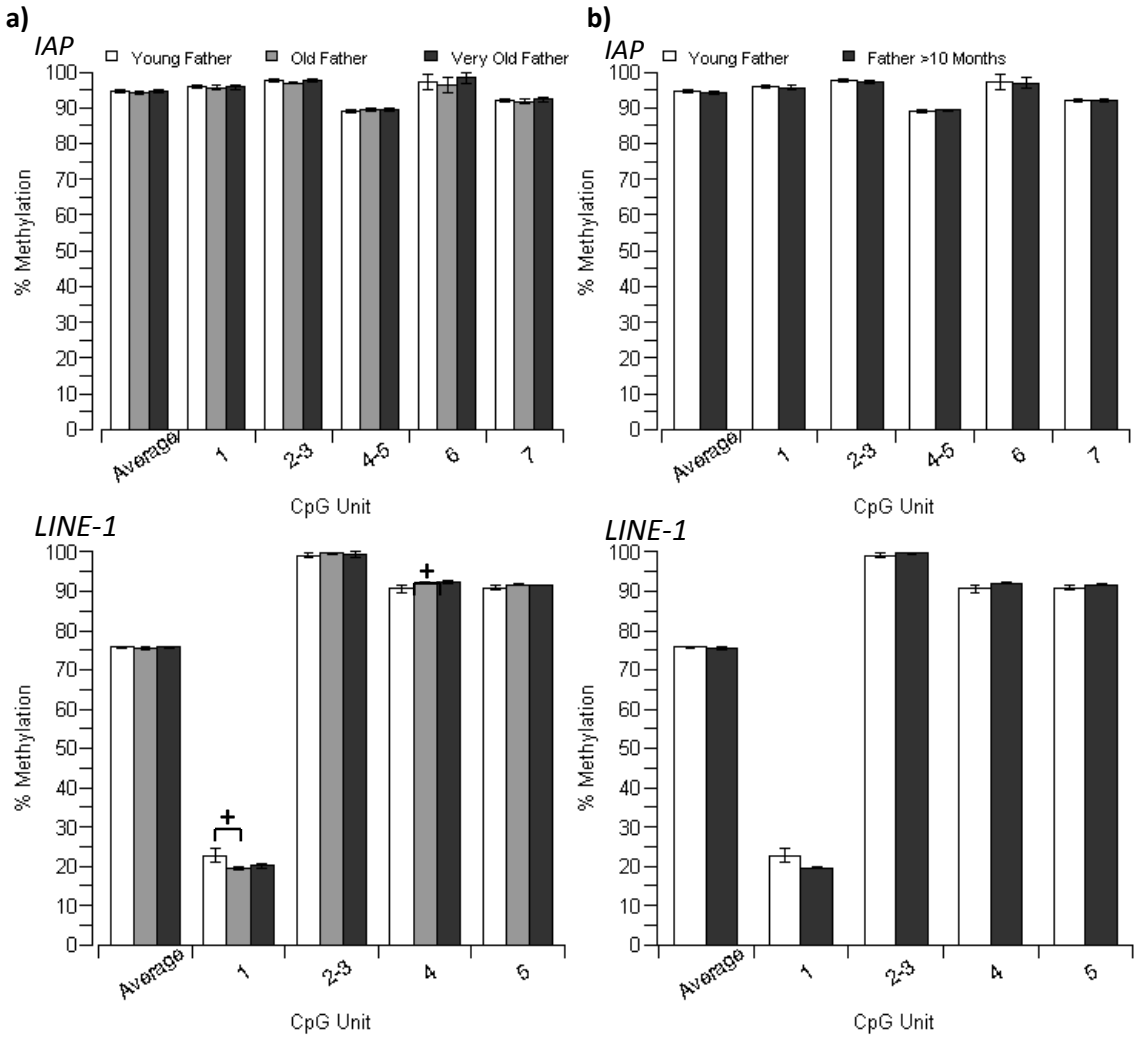
Methylation in IAP and LINE-1 elements were assayed in cerebellum samples from the offspring. There was no significant difference between groups in either IAP or LINE-1 elements between the three groups (*Figure 5.27*a). There was also no difference between the offspring of young fathers and offspring of fathers over 10 months old in IAP or LINE-1 elements (*Figure 5.27*b).

Figure 5.27- Methylation of Transposable Elements in Cerebellum Samples



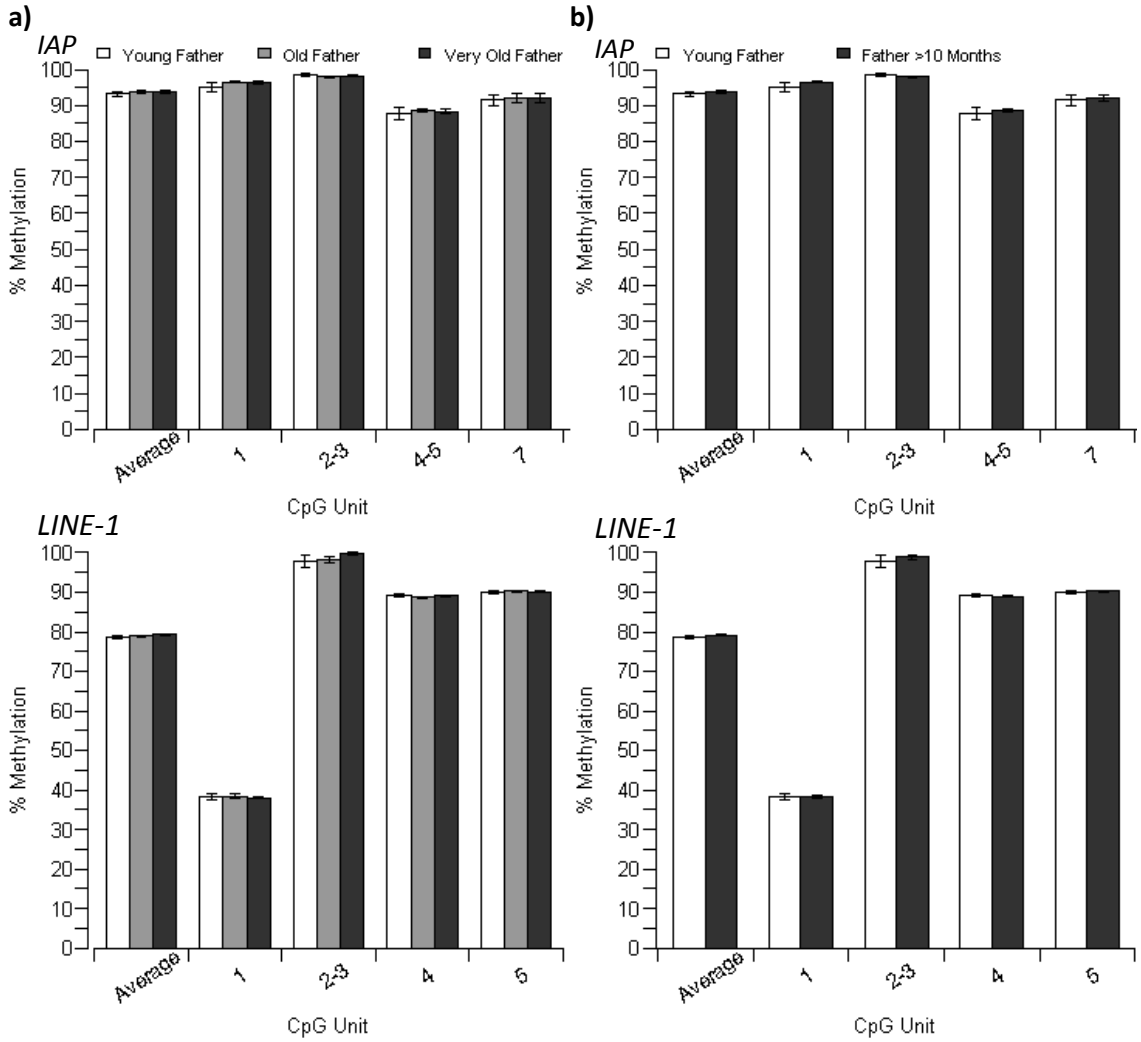
DNA from frontal cortex from the offspring of young fathers, old fathers and very old fathers were analysed for methylation in IAP and LINE1 elements. There were no significant differences between groups for either element (*Figure 5.28*a). When data from the offspring of old fathers and offspring of very old fathers was combined into offspring of fathers over 10 months, again there were no significant differences observed (*Figure 5.28*b).

Figure 5.28 - Methylation of Transposable Elements in Frontal Cortex Samples



Post mortem sperm samples from mice not related to mice used in the rest of this study were analysed for DNA methylation levels in IAP and LINE-1 elements. The results of methylation analysis in the sperm from young males (two months), old males (ten months) and very old males (twelve months) are shown in *Figure 5.29***a**. No significant differences were observed in DNA methylation across any sites in the IAP or LINE1 element assays. After combining the old male and young male data into the males over 10 months there were again no significant differences in DNA methylation between groups in IAP or LINE1 elements (Figure 5.29b).

Figure 5.29 - Methylation of Transposable Elements in Sperm Samples



5.6 Discussion

5.6.1 Summary of Results

A summary of the results from the experiments presented in this chapter is given in *Table 5.7*. An increase in global methylation, measured by LUMA, was observed in spleen and cerebellum samples with advancing paternal age but was not observed in somatic tissues from the male breeders themselves or the sperm samples from unrelated males of different ages. Initial DNA methylation analyses focussed on the cerebellum, a region of the brain that has been consistently linked with the development of autism and (to a lesser extent) psychosis (Martin and Albers 1995; Schmahmann 2000; Sparks, Friedman et al. 2002; Brambilla, Hardan et al. 2003), both of which have been shown to be epidemiologically-associated with advanced paternal age in humans. From the investigation of DMRs associated with brain

expressed imprinted genes in the cerebellum, multiple individual CpG units/sites among the other 16 amplicons investigated showed altered DNA methylation with advancing age; however only three DMRs (Gnas-Nesp, Kcnqot1 and Mcts2) showed consistent significant changes across the entire PCR region (*Figure 5.18*). Interestingly, Gnas-Nesp showed an increase in DNA methylation with increased paternal age and is expressed from the maternal allele (paternally imprinted) and Kcnq1ot1 and Mcts2 showed a decrease in methylation and are expressed from the paternal allele (maternally imprinted). This pattern is not seen across the other investigated DMRs, which were not as consistent in their DNA methylation differences. For example, Gnas XL and Zac1 are both expressed from the paternal allele (maternally imprinted) but rather than experiencing a decrease in methylation like Mcts2 and Kcnq1ot1 in DNA from the cerebellum, they both showed an increase in methylation with increasing age, although this could be due to the inconsistency in methylation pattern across the PCR region. Gnas-Nesp, Kcnq1ot1 and Mcts2 were subsequently assessed in several other tissues from the same individuals, and sperm from an independent sample of males of varying age. No significant differences were observed in spleen or hippocampus DNA, but in frontal cortex significant DNA methylation changes were observed with paternal age in all three assays. In sperm samples from males of different ages, Gnas-Nesp and Mcts2 showed an apparent highly significant decrease in DNA methylation with increasing male age. This is an interesting observation showing that the inheritance of the expressed allele could affect the change in DNA methylation differently. No significant changes in DNA methylation levels across two classes of transposable element (LINE-1 and IAP) were observed in any tissue assayed.

Table 5.7 – Summary of Results from Chapter

Results of groups compared to offspring of young fathers or young males (sperm)

Tissue	Group	Global Methylation (LUMA)	Imprinted DMRs	Transposable Elements
Spleen	**Old Father**		No differences	
	Very Old Father	Highly significant increase in methylation compared to offspring of young fathers		
	Father >10 months	Significant increase in methylation compared to offspring of young fathers		
Cerebellum	**Old Father**	Significant increase in methylation compared to offspring of young fathers	Gnas-Nesp, Kcnq1ot1 and Mcts2 showed consistent significant differences across the DMR. Nine other DMRs showed significant or trend towards significant differences compared to offspring of young fathers	No differences
	Very Old Father		Five DMRs showed sites which had significant or trend towards significant differences between groups	No differences
	Father >10 months	Trend towards increase in methylation compared to offspring of young fathers	Gnas-Nesp showed consistent methylation differences between groups across the PCR region. Eight other DMRs had significant or trend towards significant differences between groups	No differences
Hippocampus	**Old Father**	No differences	No differences	
	Very Old Father	No differences	No differences	
	Father >10 months	No differences	No differences	
Frontal Cortex	**Old Father**		Kcnqot1 showed trend towards significance in average methylation compared to young father group	No differences
	Very Old Father		Trend in methylation differences in Gnas-Nesp compared to offspring of old and young males. Kcnqot1 showed consistent significant decreased in methylation across DMR compared to old fathers. Consistently significantly lower level of methylation in Mcts2 compared to offspring of young and old fathers	No differences
	Father >10 months			No differences
Sperm	**Old Father**	No differences	Highly significant decrease in methylation across DMRs of Gnas-Nesp and Mcts2 with advanced age	No differences
	Very Old Father	No differences		No differences
	Father >10 months	No differences	Highly significant decrease in methylation across whole DMR of Gnas-Nesp and Mcts2	No differences

5.6.2 Gene Functions

The three DMRs demonstrating the most consistent differences in DNA methylation between groups in multiple tissues were Gnas-Nesp, Kcnq1ot1 and Mcts2. The Gnas-Nesp DMR regulates the expression of Nesp55 and covers exon 1A of the Gnas gene. Of note, Nesp55 knockout mice show a change of behaviour in novel environments (Plagge, Isles et al. 2005), which was also observed in behavioural analysis from Chapter 3 in the offspring of advanced paternal age mice in the holeboard task (Smith, Kember et al. 2009) (section 3.5.4). The Kcnq1ot1 DMR controls the imprinting status of Kcnq1ot1 and Cdkn1c (Engemann, Strodicke et al. 2000), regulating bidirectional gene silencing and the spread of DNA methylation on the paternally inherited chromosomes. This locus has been associated with many types of cancers and also Beckwith-Wiedemann syndrome (Biliya and Bulla 2010). Mcts2 is a relatively uncharacterised gene, resulting from the duplication of Mcts1, an X-linked gene that originated by retrotransposition into the fourth intron of H13.

5.6.3 Epidemiological Implications

As described, the epidemiological evidence linking advanced paternal age to a number of neuropsychiatric conditions, particularly autism, in humans is well-established (Frans, Sandin et al. 2008; Hultman, Sandin et al. 2010) (section 1.2.2). In Chapter 3 we reported that in the offspring of older fathers, there was reduction in exploratory and social behaviour compared to the offspring of young fathers (Smith, Kember et al. 2009). Epigenetic alterations to the regulatory regions of several imprinted loci are particularly interesting in this regard, given recent links between altered imprinting in disorders including autism (Badcock and Crespi 2006) and schizophrenia (Mill, Tang et al. 2008; Pun, Zhao et al. 2011).

The paternal age associated differences in DNA methylation could be mediated by several mechanisms. First, there is evidence for an increased prevalence of *de novo* mutations and epigenetic drift in the sperm of older males (Walter, Intano et al. 1998; Thomas, Morris et al. 2010). In humans it is estimated that there are 30 mitotic cell divisions of germ-cells before puberty then a further 23 divisions every year. There are also a further four mitotic and two meiotic divisions before the sperm are fully mature. In total, this equates to 610 chromosomal

replications by the age of 40 and 840 replications by age 50 (Crow 2000). It is possible therefore that the maintenance of imprinted marks breaks down through successive divisions leading to DNA methylation changes which are then passed on to the offspring. As there was a decrease in methylation of the three DMRs investigated in sperm as the age of the mice increased, this could be a possible explanation although additional work is required to explore this hypothesis. Many point mutations occur to convert methylated cytosine to thymine (Coulondre, Miller et al. 1978; Bird 1980), which may also alter the levels of DNA methylation across CpG-rich regions.

Compared to the fidelity of the DNA sequence which is relatively stable over many mitotic divisions, the inheritance of epigenetic marks are much less stable, with stochastic changes in DNA methylation shown to accumulate over subsequent cell divisions (Bork, Pfister et al. 2010). Previous studies have shown that there is increased intra- and inter-epigenetic variability in the sperm of older males (Flanagan, Popendikyte et al. 2006), and it is hypothesised that such changes could accumulate over the lifetime due to the effects of developmental, environmental, and stochastic factors. Since the environments of the fathers were consistent between groups in this study, our findings are unlikely to be due to an accumulation of environmentally-induced epigenetic changes, although factors such as environmental toxins are known to alter the epigenome (Yauk, Polyzos et al. 2008) and may be important in an epidemiological context. An example was reported by Anway *et al* where pregnant rats were exposed to fungicides and pesticides. The resulting male offspring presented spermatogonial cell deficits and subfertility in the F1 generation and subsequent generations. Possible causative methylation changes were mapped to 6q32 and 8q32 (Anway, Cupp et al. 2005). This provides strong evidence that epigenetic alterations can be transmitted across generation down the male germline.

Emerging evidence exists for age-related epigenetic drift in mammals, with DNA methylation being significantly correlated with age at specific loci (Bjornsson, Sigurdsson et al. 2008; Rakyan, Down et al. 2010; Bocklandt, Lin et al. 2011; Hernandez, Nalls et al. 2011; Koch, Suschek et al. 2011). Studies of MZ twins indicate that epigenetic differences between twins increase with age (Petronis, Gottesman et al. 2003; Fraga, Ballestar et al. 2005). Furthermore, aged mammalian tissue shows lower global methylation compared to younger tissue (Wilson

and Jones 1983), and several age-related cancers are caused by age-associated changes in DNA methylation (Issa, Vertino et al. 1996).

Although age-related epigenetic changes may accumulate in the germline, it has been traditionally believed that the epigenome is totally reset and re-established during gametogenesis and embryogenesis to ensure cellular totipotency (Reik, Dean et al. 2001) (section 1.6.3), preventing the meiotic transmission of such changes to subsequent generations. Recent data, however, indicates that transgenerational epigenetic inheritance may be possible (Daxinger and Whitelaw 2010; de Boer, Ramos et al. 2010), and because several imprinted regions appear to escape germ-line epigenetic reprogramming (Popp, Dean et al. 2010), these elements could provide a vehicle for the transmission of the paternal age effect, enabling the epigenetic changes accumulating during the longer development of the sperm cells in the father to be transferred to the offspring. A recent study has shown that DNA methylation in newborn babies is correlated with parental age at over 1000 genes, including several known imprinted loci (Adkins, Thomas et al. 2011). As we detect some evidence that the global methylation level is also increased with increasing father's age, albeit only detectable in certain tissues, the changes to DNA methylation could be more widespread than just imprinted loci, which account for a small proportion of the genome.

5.6.4 Limitations

There are several limitations to this study. We were unable to examine sperm DNA from the fathers of the offspring used in the rest of the study, so we are unable to conclude that epigenetic changes observed in the offspring are reflected by changes occurring in their fathers' sperm. Of note, however, we observe DNA methylation differences in two DMRs in sperm from age matched male mice from another source, providing some plausibility to the theory that the paternal age effect could be mediated by epigenetic inheritance via the male germline. These results should, however, be treated with some caution given that we were unable to obtain 100% pure populations of sperm cells. Only three ages of fathers were analysed; it would be interesting to examine DNA methylation differences in the offspring of even older fathers to see if more dramatic changes are observed. Our study only focused on specific genomic features (i.e. DMRs associated with several brain-expressed imprinted genes) and so we cannot say anything about epigenetic changes in other regions of the genome. Only

DNA methylation was assessed at these loci, and it is plausible that other epigenetic mechanisms (i.e. histone modifications and non-coding RNAs) may also be altered in this context. The IAP and LINE-1 element assays were designed to cover many genomic locations which share a consensus sequence. However, due to variations in the genomic sequences in these elements, the assays do not cover all instances of these features across the genome. As these assays provide an estimate of average DNA methylation across the genome, it is possible we are missing differences in gene-associated elements.

5.6.5 Conclusions

In summary, we used a rodent model of advanced paternal age, previously shown to exhibit behavioural phenotypes related to deficits observed in human neuropsychiatric disease, to identify changes in DNA methylation across DMRs associated with several brain-expressed imprinted genes in the offspring of older fathers as well as global methylation and methylation at transposable elements. Our results provide a potential mechanism for the paternal age effect seen in neuropsychiatric disorders such as autism and schizophrenia, and suggest that epigenetic alterations to the expression of imprinted loci could be involved in the aetiology of these disorders.

Chapter 6 - Genome Wide Analysis of Gene Expression in a Mouse Model of Advanced Paternal Age

6.1 Abstract

Previous studies of gene expression in autism, schizophrenia and bipolar disorder have identified widespread transcriptomic changes associated with these diseases. In autism, for example, evidence suggests that genes involved in cell signalling pathways are differentially expressed in the brain. Expression changes have also been observed in studies of aging, especially in genes associated with inflammatory response and DNA repair. In this chapter, I examined frontal cortex gene expression differences between the offspring of young and old male mice using Illumina Mouse Ref8 V2 BeadChips, targeting approximately 25,600 RefSeq transcripts. Numerous genes with nominally significant differential expression between groups were identified, although the overall magnitude of differences observed was small. Interestingly, pathway analysis on these differentially expressed genes shows enrichment for genes associated with cell signalling (as in autism) and inflammatory response and DNA repair (previously associated with aging).

6.2 Introduction

6.2.1 Gene Expression and Paternal Age

As discussed in relation to DNA methylation (section 5.2.4), the transcriptional profile of a cell is characterised by significant age-related changes. Expression changes associated with aging in the brain have been investigated largely in terms of neurodegeneration and Alzheimer's disease (Colangelo, Schurr et al. 2002), although aging in general has been associated with widespread transcriptional changes. In a mouse study of aging, for example, brain samples from cerebellum and neocortex displayed increased expression of genes related to inflammatory and stress response and decreased expression of genes associated with growth and trophic factors, protein turnover, DNA synthesis and repair and neurotransmission (Lee, Weindruch et al. 2000). Age-associated changes are observed in numerous tissue- and cell-types, including the germ-line. Of relevance to my investigations of paternal age, an expression study of spermatocytes from rats of different ages identified over 2800 loci that are differentially expressed in spermatocytes from older males (18 months) compared to spermatocytes from young males (4 months); of note, many genes associated with base excision repair, nucleotide excision repair, mismatch repair and double strand break repair were altered in spermatocytes from older males (Paul, Nagano et al. 2011). Another study by Kokkinaki *et al* investigated RNA from the spermatogonial stem cells of mice of four different

ages (6 days, 21 days, 60 days and 8 months old) using microarrays. 2819 genes showed differential expression between the age groups ($p<0.05$, fold change >2) including genes previously identified in gene expression studies of aging in stem cells (Kokkinaki, Lee et al. 2010). Pathway analysis on these genes highlighted an enrichment of genes involved in DNA repair and oxidative stress. Given the known increase in DNA damage in the spermatozoa of older males (section 1.4.1), this is an interesting finding. To date, however, only one study has looked at gene expression changes in association with paternal age (Alter, Kharkar et al. 2011). In this study of paternal age and autism, a decrease in the overall variance in gene expression was observed in the offspring of older fathers, in addition to a down regulation of genes involved in gene transcription (Alter, Kharkar et al. 2011).

6.2.2 RNA Expression and Psychiatric Disorders

Changes in gene expression have been observed in many psychiatric disorders, including those associated with advanced paternal age. Cerebellum samples from autism cases showed 15 up-regulated and 11 down-regulated genes compared to control cerebellum samples including members of the glutamate system which were significantly increased in cases (Purcell, Jeon et al. 2001). Reelin (*RELN*) has been shown to be down-regulated in autistic patients compared to controls (Fatemi, Snow et al. 2005), an interesting observation given that polymorphisms in RELN have been previously associated with autism (Zhang, Liu et al. 2002). MZ twins discordant for autism severity displayed multiple gene expression differences including the serotonin transporter gene (*5HTT*), tumour necrosis factor (*TNF*) and other cytokine genes (Hu, Frank et al. 2006). Cell-line samples from autism cases and controls showed 30 up-regulated and 18 down-regulated genes in cases compared to controls, including genes involved in metabolism and the regulation of metabolic pathways, cell growth and maintenance and cell proliferation (Baron, Liu et al. 2006). In blood samples from autistic children and controls, 55 genes showed differential expression with the top pathway showing expression differences between groups being the natural killer cell signalling pathway (Gregg, Lit et al. 2008). Expression of *Sp1*, *CD38*, *ITGB3*, *MAOA*, *MECP2*, *OXTR*, and *PTEN* all showed significant elevation in expression and *GABRB3*, *RELN*, and *HTR2A* showed significant reductions in expression in a study of brain tissue from autism patients compared to controls (Thanseem, Anitha et al. 2011). Finally, in Voineagu *et al*'s study of post-mortem brain samples, 444 genes were shown to be differentially expressed between autism cases and controls in cerebral cortex samples. In this data they found common transcriptomic signatures

across cases after modulating the data into co-expression networks (Voineagu, Wang et al. 2011)

Numerous transcriptomic studies have been performed in schizophrenia, and multiple genes have been found to display altered expression in cases compared to controls. These include *DTNBP1*, which has been shown to be decreased in post-mortem brain samples from schizophrenia patients compared to controls (Weickert, Straub et al. 2004). *NRG1* expression is increased (Hashimoto, Straub et al. 2004) and *RGS4* is significantly reduced in schizophrenia patients (Mirnics, Middleton et al. 2001). Expression of *CHRNA7* was reduced in multiple brain regions in schizophrenia patients (Freedman, Hall et al. 1995; Court, Spurden et al. 1999; Guan, Zhang et al. 1999; Marutle, Zhang et al. 2001) and *Akt1* expression in reduced in blood samples from schizophrenia patients (Emamian, Hall et al. 2004). Expression of the *5HT2A* receptor has been shown to be reduced in schizophrenia patients (Harrison 1999) as well as a reduction in expression of glutamate receptors such as *GluR1* and *GluR2* (Eastwood, McDonald et al. 1995; Eastwood, Kerwin et al. 1997). Expression of *TGF1-β1* has been observed to be higher and *IL-4* lower in schizophrenia patients than controls (Kim, Myint et al. 2004). Finally, reductions in oligodendrocyte-related and myelin-related genes (Tkachev, Mimmack et al. 2003), reduced expression of mitochondria related genes (Altar, Jurata et al. 2005), and lowered transcription of GABAergic neuron markers (Kuromitsu, Yokoi et al. 2001; Pongrac, Middleton et al. 2002) have all also been observed in post mortem brain samples of schizophrenia patients compared to controls.

BD has also been associated with altered transcription at multiple loci. Up regulation of *5HTT* and *NFkB2* (Sun, Zhang et al. 2001) and *NFkB* (Spiliotaki, Salpeas et al. 2006) were observed in BD patients. In genome-wide studies several genes demonstrate altered expression in bipolar patients compared to controls (Bezchlibnyk, Wang et al. 2001; Iwamoto, Kakiuchi et al. 2004; Jurata, Bukhman et al. 2004) including down regulation of neurotransmitter related genes (Iwamoto, Kakiuchi et al. 2004), increased expression of stress response genes (Iwamoto, Kakiuchi et al. 2004; Jurata, Bukhman et al. 2004), increased expression of chaperones (Jurata, Bukhman et al. 2004), decreased expression of mitochondrial related genes (Konradi, Eaton et al. 2004; Sun, Wang et al. 2006; Vawter, Tomita et al. 2006) and changes to expression in GABAergic genes (Choudary, Molnar et al.

2005) and glutamate receptor genes (Iwamoto, Kakiuchi et al. 2004; Nakatani, Hattori et al. 2006).

Many overlapping genes and gene pathways have been identified in schizophrenia and BD. *TGF1-β1* displays altered expression in both disorders (Bezchlibnyk, Wang et al. 2001; Kim, Myint et al. 2004) although it is reduced in BD patients and increased in schizophrenia patients. Down regulation of oligodendrocyte related genes have been observed in genome wide studies on post mortem brain samples from both schizophrenia and BD patients (Tkachev, Mimmack et al. 2003). Finally, down regulation of mitochondrial genes has been observed in studies of both BD (Konradi, Eaton et al. 2004) and schizophrenia (Altar, Jurata et al. 2005).

6.2.3 Relationship between DNA Methylation and Gene Expression

As discussed in Chapter 1, DNA methylation at gene promoters is often inversely correlated with the level of gene expression (Jaenisch and Bird 2003). Highly methylated promoter regions bind methyl-CpG-binding domain proteins (MBDs), which can act as adaptors between methylated DNA and histone-modifying enzymes and chromatin remodelling proteins. MECPs recruit histone-modifying enzymes to regions of methylated DNA (Hendrich and Tweedie 2003). The resulting closed and compacted chromatin structure (heterochromatin) blocks the binding of transcription factors and thus represses gene transcription (Jenuwein and Allis 2001). Conversely, low DNA methylation at gene promoters leads to an open chromatin structure (heterochromatin) that is not associated with the recruitment of histone-modifying enzymes. DNA methylation levels are often variable by tissue allowing for tissue specific transcriptomic patterns to be established during development. Recent data, including from our group (Davies, Volta et al. Under Review), suggests that non-promoter CpG islands may be even more important for establishing tissue-specific gene expression than promoter CpG islands, especially those located intergenically (located in areas of no genes) or intragenically (located within the gene-body). As discussed in section 1.6.2, allele-specific DNA methylation at key DMRs and ICRs plays a crucial role in establishing the monoallelic expression patterns associated with imprinted genes. DNA

methylation is dynamically regulated and variable by developmental stage and age, allowing coordinated transcription of genes required during different developmental stages (Paulsen and Ferguson-Smith 2001). Given my findings related to epigenetic changes in the context of advanced paternal age in Chapter 5, it is plausible that wider alterations to the transcriptome may be detected in the offspring of older fathers.

6.3 Aims

The aim of this chapter is to identify genes and transcriptomic pathways that are differentially expressed in mouse between the offspring of young fathers and offspring of old fathers using genome-wide gene expression microarrays on RNA extracted from a region of the brain (frontal cortex) previously implicated in the neuropsychiatric disorders epidemiologically-associated with advanced paternal age in humans.

6.4 Methods

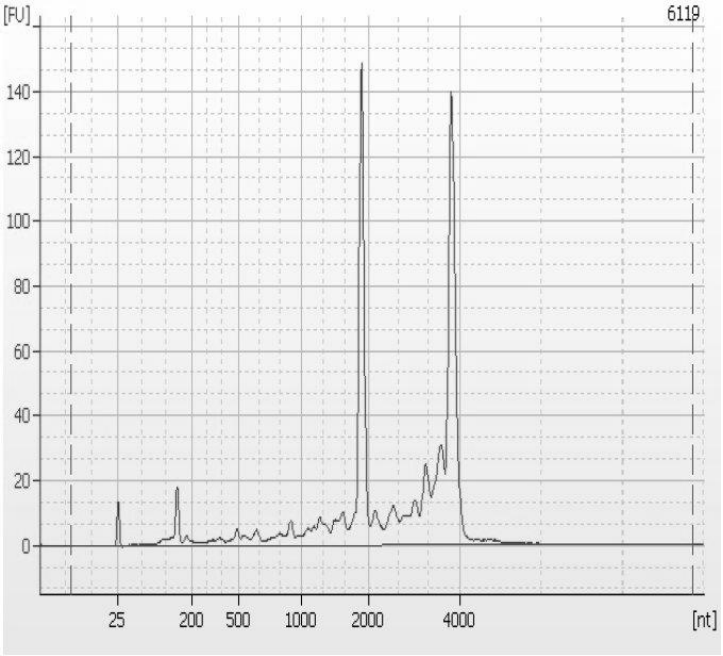
6.4.1 RNA Extraction and Sample Selection

RNA was extracted from frontal cortex samples from the offspring of young fathers and offspring of old fathers. The frontal cortex was chosen for expression analysis due to its role language and social and emotional functions (Stuss and Knight 2002). Frontal cortex has been shown to develop abnormally in children with autism (Kemper and Bauman 1998; Carper, Moses et al. 2002), and alterations to expression and protein levels observed in frontal cortex from schizophrenia and BD patients (Johnston-Wilson, Sims et al. 2000). 36 samples were selected (16 offspring of old fathers, 16 offspring of young fathers) to represent all families from the offspring of young fathers (two months old) and old fathers (ten months old). RNA from frontal cortex was extracted using Qiagen AllPrep RNA/DNA as described in detail in section 2.2.2. Stringent quality control was carried out on the RNA so that all samples had a 260/280 ratio of ≥ 1.9 and a clear 23S and 16S fragment on a 1.5% agarose gel with minimal degradation (*Figure 2.2*). RNA integrity was further assessed by running samples on the Agilent Bioanalyzer 2100 using the RNA 6000 Pico Assay (*Figure 6.1*). The RNA integrity number (RIN) is a value provided by the Agilent Bioanalyser software to estimate the integrity of total RNA samples. This assesses the entire electrophoretic trace of the RNA sample

including the presence or absence of degradation products. A RIN number of 10 indicates a hypothetically 'perfect' RNA sample; a RIN number of 5 indicates partially degraded RNA and a RIN of 1 totally degraded RNA. All samples used in this experiment reached a RIN number of greater than 7 except one sample (PA31) which had a RIN number of 6.1. This sample was still included in the Illumina array experiment, but was flagged for caution in the subsequent analyses.

Figure 6.1 – Example Electropherograms from Bioanalyzer Agilent RNA Pico Labchip

Sample PA02, RIN Score 8.4. Intact total RNA is represented by two strong ribosomal peaks representing 28S and 18S species.



6.4.2 IlluminaTotalPrep-96 RNA Amplification Kit

The IlluminaTotalPrep-96 RNA amplification kit was used to generate biotinylated amplified RNA for hybridization with Illumina arrays. 300ng of quantified RNA was used for each sample and the protocol carried out as in the manufacturer's instructions (*Figure 6.2*).

Figure 6.2 - Illumina® TotalPrep™-96 RNA Amplification Procedure Overview

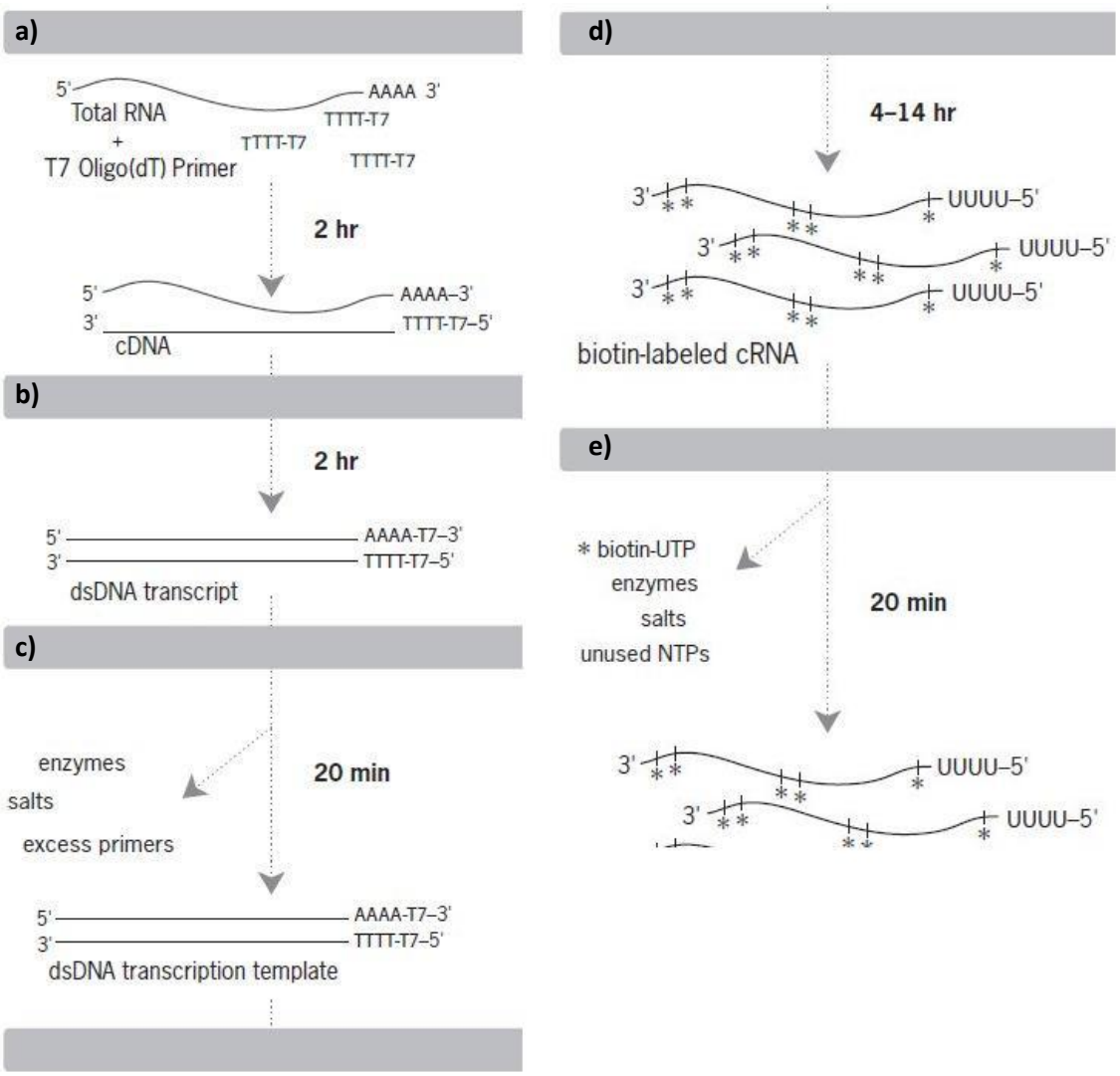
a) Reverse Transcription to Synthesize First Strand cDNA

b) Second Strand cDNA Synthesis

c) cDNA Purification

d) In Vitro Transcription to Synthesize Biotin-labelled cRNA

e) cRNA Purification



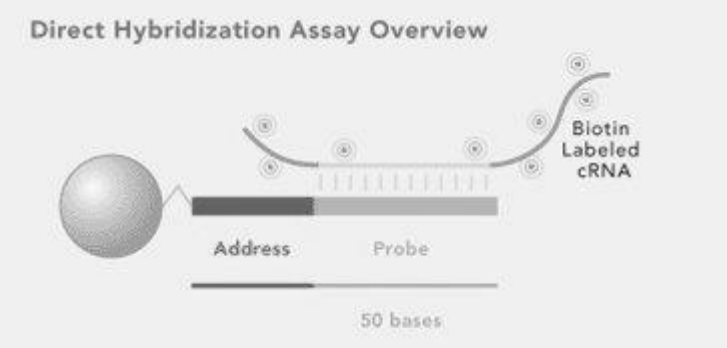
6.4.3 Whole-Genome Gene Expression Direct Hybridization

RNA samples were analysed using Illumina Mouse Ref8 V2 BeadChips. The Illumina Mouse Ref8 V2 array targets approximately 25,600 RefSeq transcripts and over 19,100 unique genes derived from the NCBI database (build 36/mm8) supplemented with probes derived from the

Mouse Exonic Evidence Based Oligonucliotide (MEEBO) set as well as exemplar protein-coding sequences described in the RIKEN FANTOM2 database. Eight samples were run in parallel on each chip and samples randomised and distributed across the arrays to ensure a balance of samples from both groups. Whole-Genome Gene Expression Direct Hybridization Assay was carried out as in manufacturer's instructions (*Figure 6.3*), and BeadChips were scanned using the Illumina iScan System.

Figure 6.3 – Direct Hybridization Assay

A 50 bases gene-specific probe linked to a short address. The probe is hybridized to labelled cDNA derived from total RNA



6.4.4 Data Analysis

6.4.4.1 LUMI

Data was normalized and initially analysed using the "lumi" R package (Du, Kibbe et al. 2007; Du, Kibbe et al. 2008; Lin, Du et al. 2008) (R scripts used for the lumi package are shown in *Appendix 3*). Briefly, lumi first log2 transforms all intensity data then implements an optimised normalization procedure. Pre-normalized expression values and data description are shown in *Figure 6.4* and *Table 6.1*. As expected, the mean signal intensity, standard deviation and the detection rates were somewhat variable across samples, but there were no obvious extreme outliers indicative of poor data quality. The sample with the slightly lower RIN score (PA31), showed no obvious abnormalities in terms of raw data distribution

Figure 6.4 - Prenormalization Plots of Array Intensities

a) Density Plot of Intensities

b) Cumulative Distribution Function

c) Boxplot of Microarray Intensities

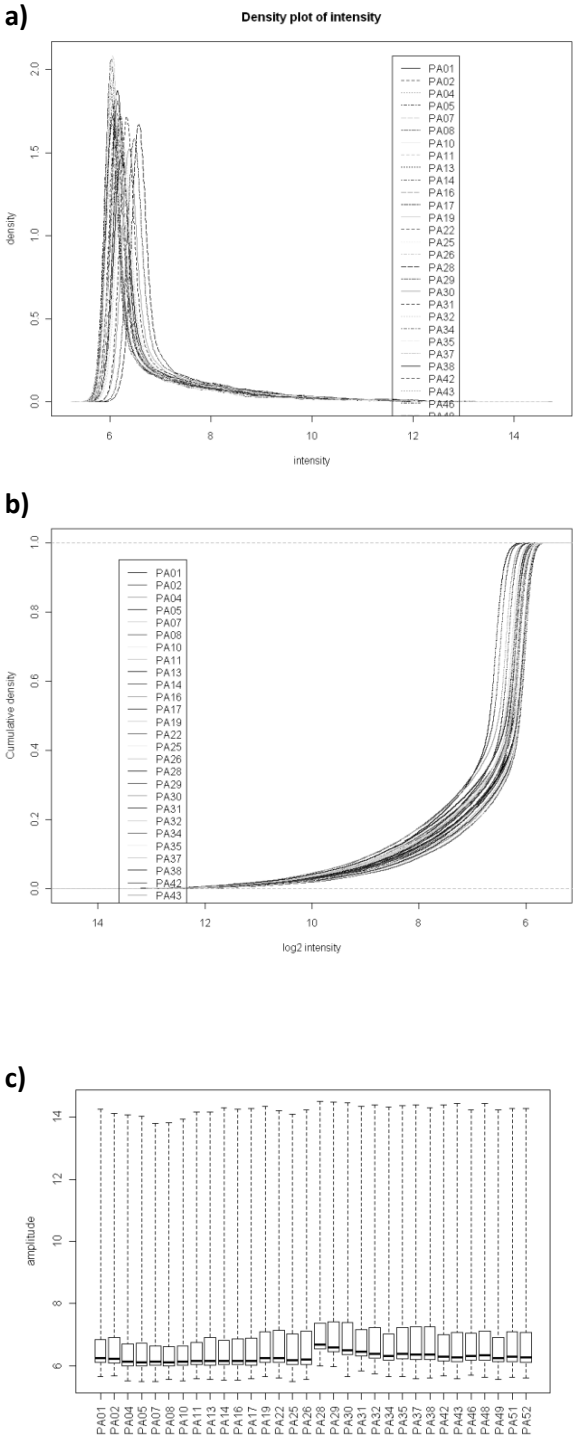


Table 6.1 - Prenormalization Values of Illumina Microarray Data

Sample	PA01	PA02	PA04	PA05	PA07	PA08	PA10	PA11	PA13	PA14	PA16
mean	6.696	6.717	6.573	6.57	6.555	6.532	6.549	6.619	6.678	6.651	6.668
standard deviation	1.047	1.1	1.035	1.049	0.9928	0.9911	1.006	1.068	1.147	1.102	1.125
distance to sample mean	27.41	26.1	43.45	43.37	49.1	51.36	49.13	34.05	23.63	27.66	24.95

Sample	PA17	PA19	PA22	PA25	PA26	PA28	PA29	PA30	PA31	PA32	PA34
mean	6.678	6.803	6.832	6.741	6.78	7.158	7.141	7.075	6.945	6.946	6.805
standard deviation	1.143	1.198	1.209	1.189	1.238	1.109	1.175	1.214	1.119	1.2	1.121
distance to sample mean	24.1	21.67	27.42	24.84	26.24	62.13	60.04	51.58	29.65	33.15	18.06

Sample	PA35	PA37	PA38	PA42	PA43	PA46	PA48	PA49	PA51	PA52
mean	6.926	6.938	6.929	6.804	6.818	6.822	6.86	6.735	6.825	6.801
standard deviation	1.196	1.218	1.222	1.132	1.185	1.133	1.171	1.093	1.168	1.171
distance to sample mean	32.33	33.29	34.68	20.48	22.66	18.32	22.58	20.6	21.53	16.95

To normalise across arrays, variance stabilizing transformation was performed; this is a standard data transformation which stabilizes the variance within each microarray (Lin, Du et al. 2008). The next step of normalization employed robust spleen normalization (RSN), which is unique to the lumi package and combines features of quantile normalization and loess normalization. This combination allows the preservation of small differences between groups and ensures the rank order of genes remains unchanged across samples. RSN calibrates each array to one reference array (in this case PA01). First the data is quantile normalised then the fold change of a gene is measured based on the quantile normalised data. Postnormalization array data is shown in *Figure 6.5* and *Table 6.2*. As expected, after normalization the means, standard deviations and detection rates were much more uniform across the samples. Results from the lumi analysis are described in section 6.5.1.1.

Figure 6.5 - Postnormalization Plots of Array Intensities

a) Density Plot of Intensities

b) Cumulative Distribution Function

c) Boxplot of Microarray Intensities

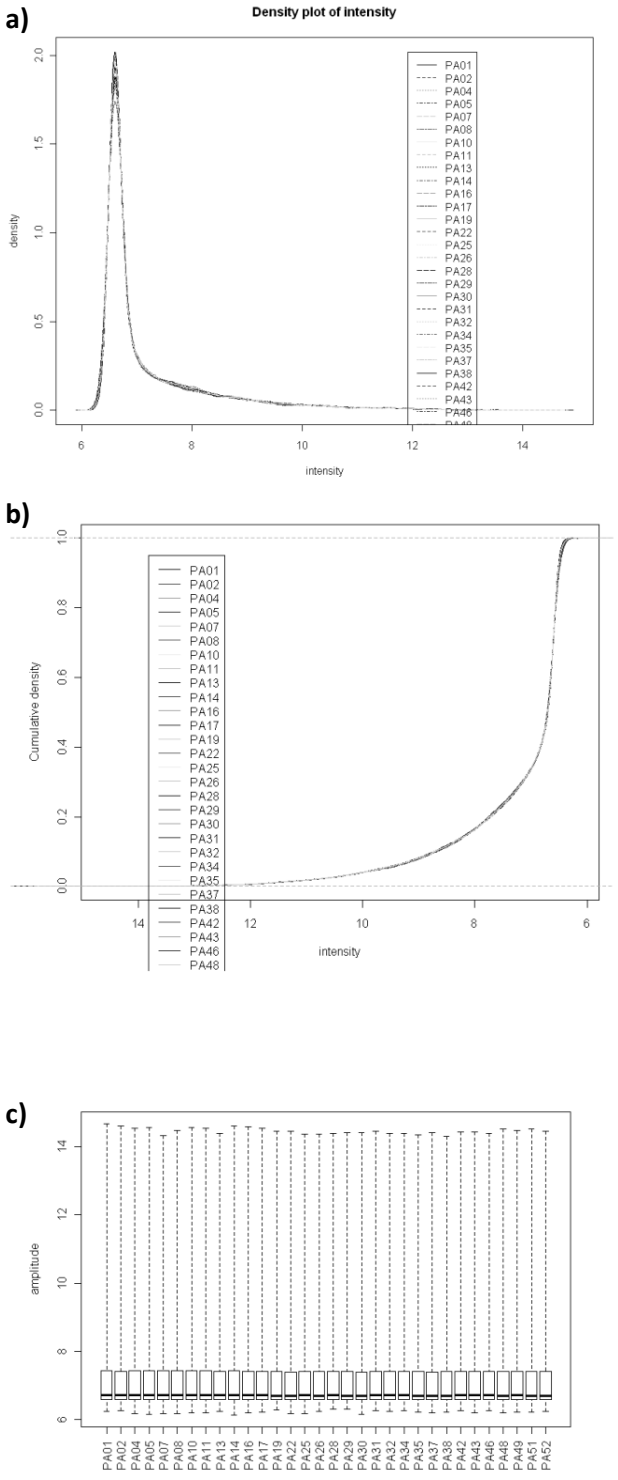


Table 6.2 - Postnormalization Values of Illumina Microarray Data

Sample	PA01	PA02	PA04	PA05	PA07	PA08	PA10	PA11	PA13	PA14	PA16
mean	7.05	7.05	7.049	7.049	7.05	7.048	7.05	7.049	7.05	7.048	7.049
standard deviation	1.024	1.024	1.024	1.024	1.024	1.025	1.024	1.025	1.024	1.025	1.024
distance to sample mean	16.07	20.69	18.55	19.99	18.21	18.08	19.46	15.69	13.27	14.2	13.64

Sample	PA17	PA19	PA22	PA25	PA26	PA28	PA29	PA30	PA31	PA32	PA34
mean	7.049	7.05	7.049	7.05	7.049	7.05	7.05	7.05	7.049	7.05	7.05
standard deviation	1.025	1.023	1.025	1.024	1.025	1.024	1.024	1.024	1.024	1.024	1.023
distance to sample mean	14.24	14.77	18.33	18.13	13.65	14.86	14.44	13.17	13.27	13.84	15.52

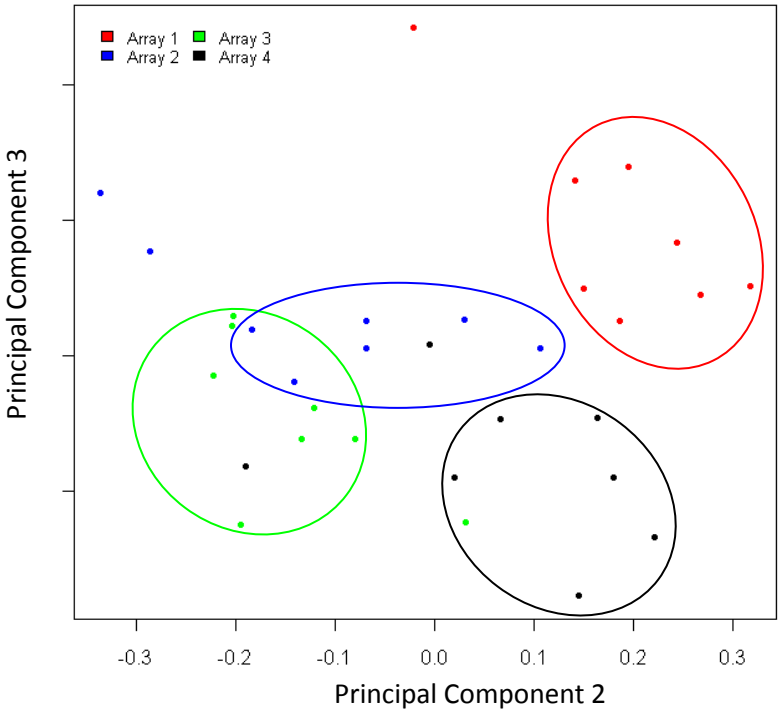
Sample	PA35	PA37	PA38	PA42	PA43	PA46	PA48	PA49	PA51	PA52
mean	7.049	7.049	7.049	7.05	7.049	7.05	7.049	7.049	7.049	7.049
standard deviation	1.024	1.024	1.024	1.024	1.024	1.024	1.025	1.024	1.025	1.025
distance to sample mean	16.69	12.14	16.57	17.61	17.4	14.77	15.14	15.48	16.04	12.46

6.4.4.2 Correcting for Array Batch Effects using ComBat

After normalization, principal components analysis (PCA) indicated that the data was potentially characterized by a strong batch effect (*Figure 6.6*), as widely reported for Illumina gene expression arrays. As my hybridization experiment was designed to ensure that samples were randomly assigned to arrays and balanced for the number of samples from each group on each array, it is unlikely that this batch affect will mask any true expression differences (or create false positives) between groups. However, batch affects were adjusted for using the R ComBat package. This package uses parametric and nonparametric empirical Bayes frameworks for adjusting data for batch effects (Johnson, Li et al. 2007). After ComBat correction had been applied, the lumi package analysis was rerun (section 6.5.1.2).

Figure 6.6 – Results of Principal Components Analysis

Plot of 2nd against 3rd principal component coloured by array the sample was run on. Samples show a strong batch effect as indicated by clustering of samples.



6.4.4.3 Weighted Gene Co-Expression Network Analysis

Weighted Gene Co-Expression Network Analysis (WGCNA) (in R) looks for relationships between measured transcripts that are detected by pair-wise correlations between genes

(Langfelder and Horvath 2008) (R scripts for WGCNA is shown in *Appendix 3*). WGCNA has been previously employed to identify clusters (or modules) of highly co-expressed genes from large transcriptomic datasets. WGCNA starts from the level of thousands of genes, identifies phenotypically interesting gene modules, and uses intramodular connectivity to identify key genes in functional pathways for further validation.

WGCNA was used to construct sample dendrogram and trait heat-maps across the expression dataset, identify co-expression networks and detect modules of co-expressed genes. Constructing a weighted gene network entails the choice of the soft thresholding power to which co-expression similarity is raised to calculate adjacency. The soft-thresholding power was selected by plotting the scale-free fit index as a function of the soft-thresholding power. The power 6 was chosen to construct the network which is the lowest power for which the scale-free topology t index reaches 0.90. The gene network and identifying modules were then run using a minimum module size of 100 genes. After the networks were constructed, the data was split by WGCNA into 22 different modules.

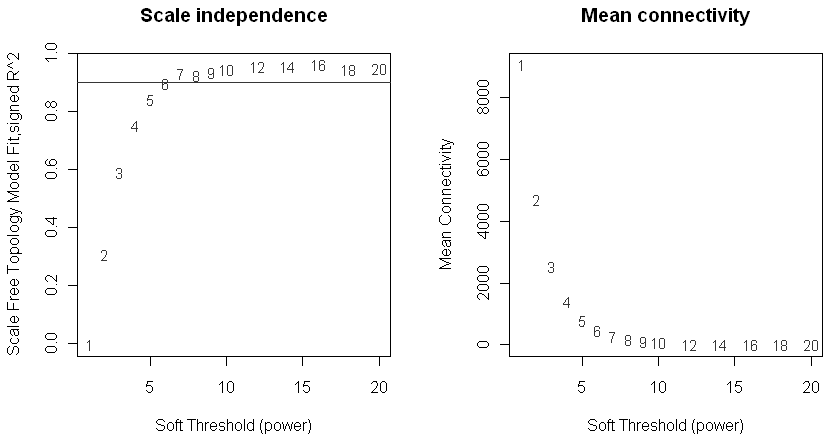


Figure 6.7 - Analysis of Network Topology for Various Soft-Thresholding Powers

The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis).

6.4.4.4 One Sample Student's T-test

We next compared the expression of genes encompassed by CNVS (from Chapter 4). As most of the CNVs identified in our analyses are rare events, generally private to specific individual animals, a standard t-test is not appropriate to assess the effect of a CNV on expression of the genes they span. In this case, a one sample t-test was used within the R statistical environment to compare locus-specific gene expression in CNV carriers (designated mu) against the average expression for that gene across all individuals.

6.5 Results

6.5.1 LUMI

6.5.1.1 Pre-Batch Correction

Probes for 18,097 genes were detected at a high enough level to be considered expressed by lumi and so were used for further analysis. Comparing frontal cortex gene expression in the offspring of old versus young fathers, 371 genes reached an uncorrected p-value of less than 0.05. Genes with a p-value <0.001 are shown in *Table 6.3* and the top ten most significantly differentially expressed genes are shown graphically in *Figure 6.8*. The false discovery rate (FDR) was used to correct for multiple comparisons and generate q-values. The FDR approach is optimised by using characteristics of the p-value distribution to produce a list of q-values. No genes in this initial analysis on non batch-corrected data had an FDR below 0.1.

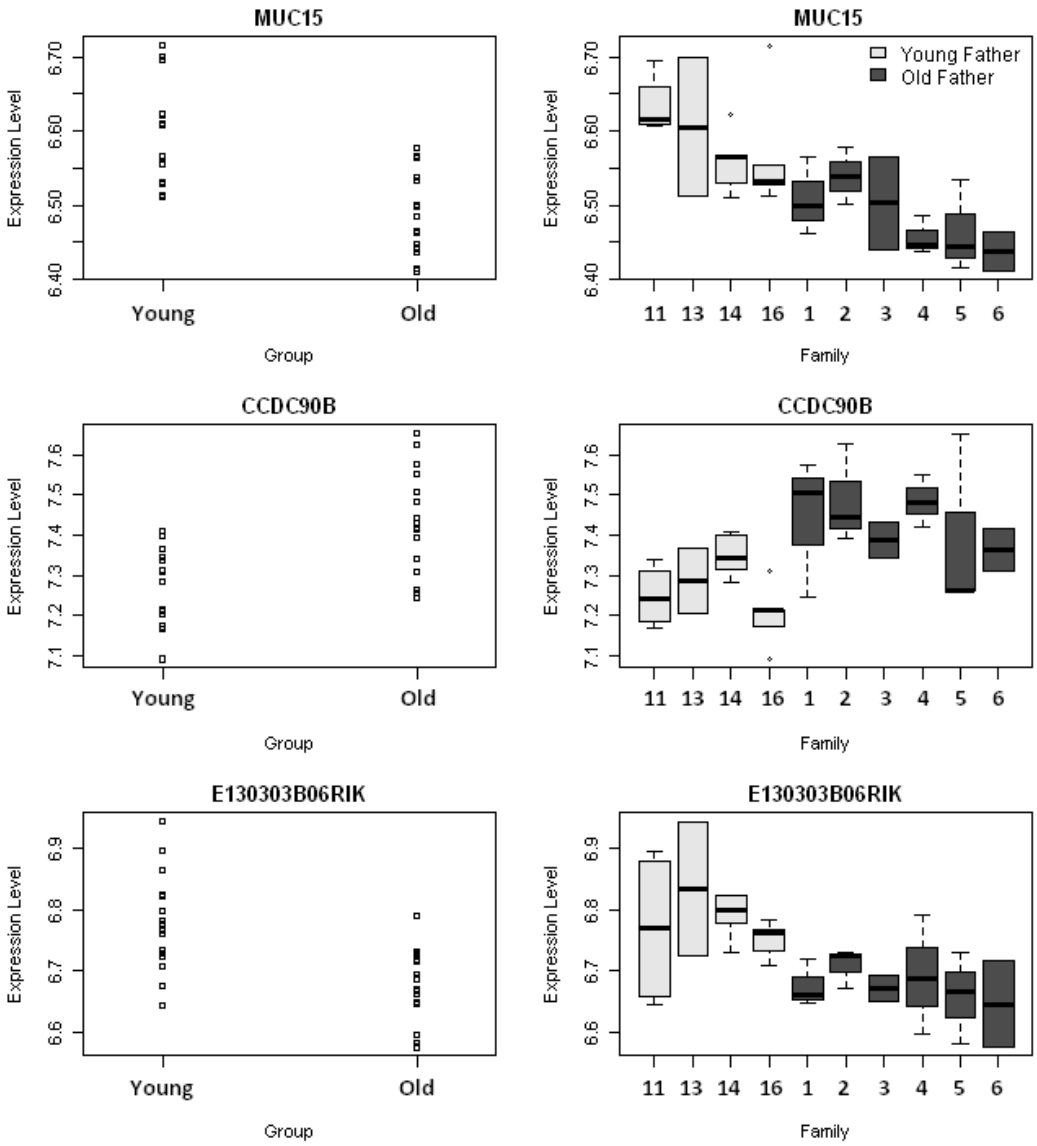
Table 6.3 - Top Differentially Expressed Genes from lumi Package Pre-Batch Correction

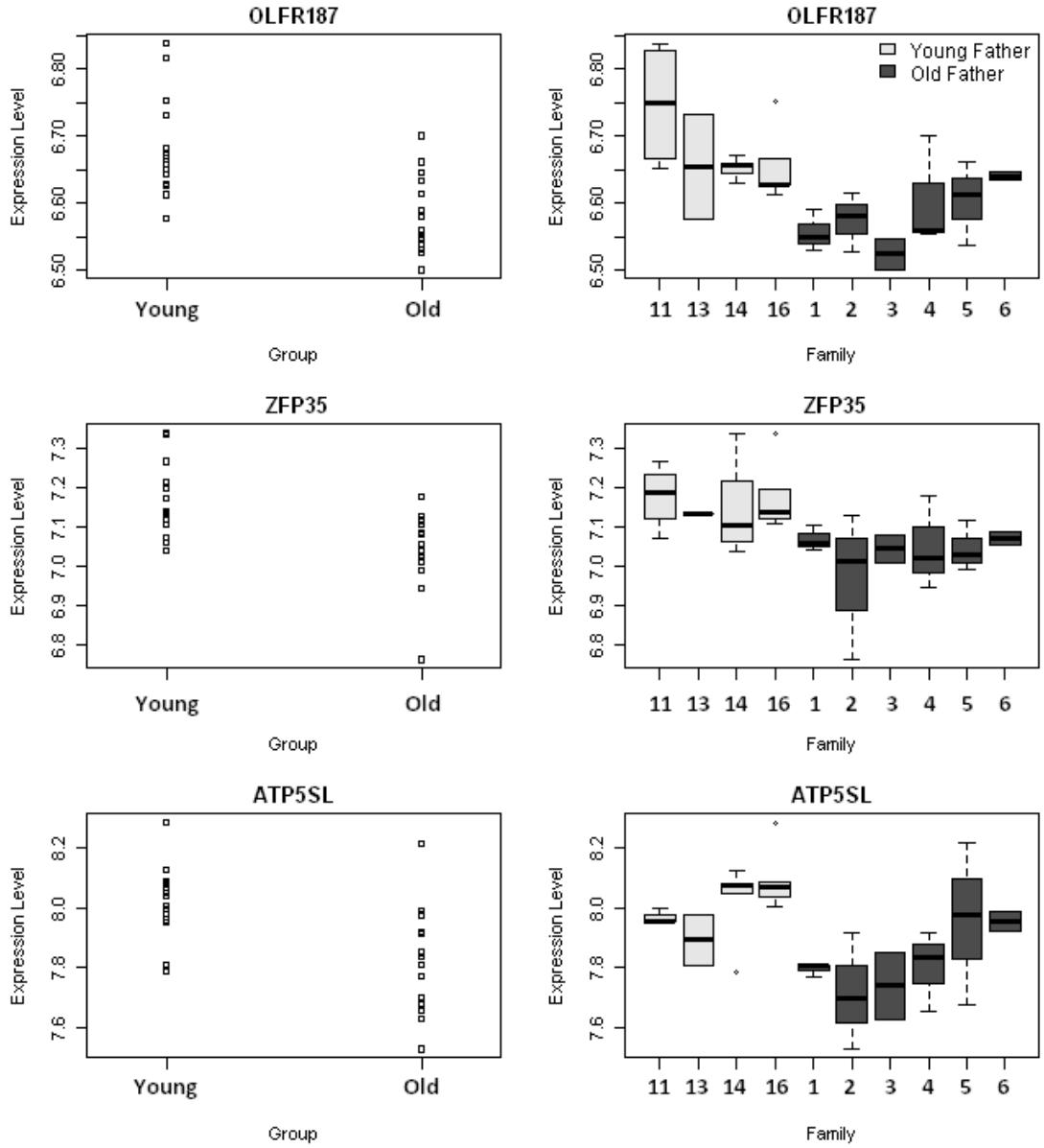
Genes which showed an uncorrected p-value of ≤ 0.001

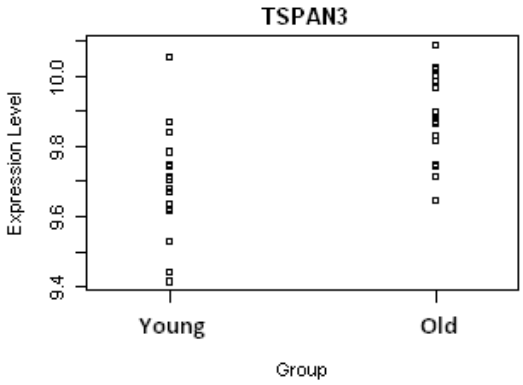
Gene	Average	Average Young	Average Old	t	Uncorrected p-Value
MUC15	6.54	6.59	6.49	4.39	1.0E-04
CCDC90B	7.35	7.27	7.43	-4.25	1.6E-04
E130303B06RIK	6.73	6.78	6.68	4.02	3.0E-04
OLFR187	6.63	6.68	6.58	4.02	3.1E-04
ZFP35	7.10	7.16	7.04	3.96	3.7E-04
ATP5SL	7.92	8.02	7.82	3.89	4.5E-04
TSPAN3	9.79	9.69	9.88	-3.86	4.8E-04
OLFR92	6.58	6.54	6.61	-3.85	0.001
OLFR1366	6.52	6.57	6.48	3.76	0.001
E030049G20RIK	6.61	6.64	6.57	3.57	0.001
PSG25	6.64	6.67	6.60	3.56	0.001
LARP4	6.55	6.58	6.52	3.46	0.001
IER5L	7.77	7.84	7.69	3.46	0.001

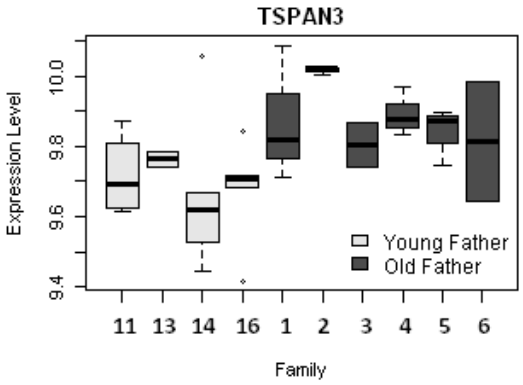
Figure 6.8 – Boxplot of Top 10 Differentially Expressed Genes Pre-Batch Correction

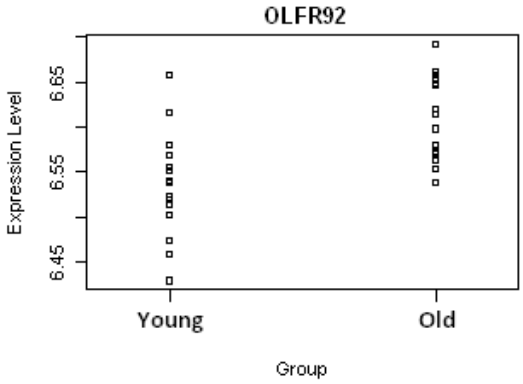
Top 10 differentially expressed genes from lumi analysis. All genes show strip charts of offspring of young fathers and offspring of old fathers as well as box plot of expression level by family. Black line in box plots shows median value, edges of boxes show upper and lower quartiles and whiskers show the upper and lower extreme values.

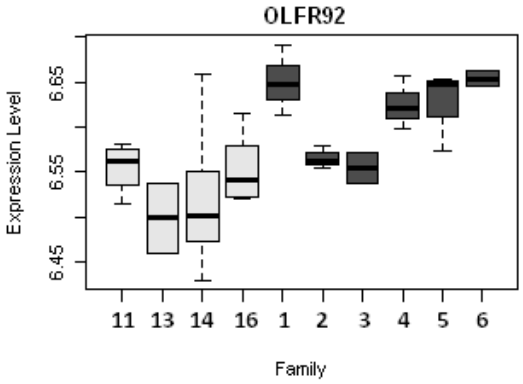


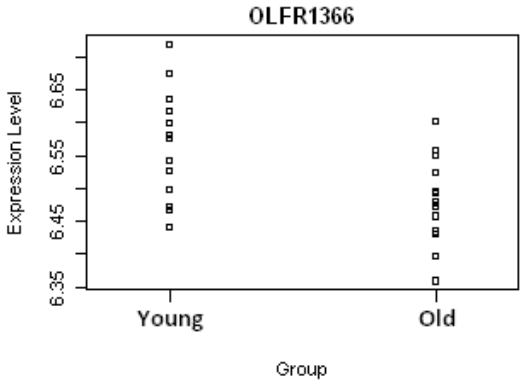

OLFR187
Expression Level
6.50
6.60
6.70
6.80
Young
Old
Group
OLFR187
Young Father
Old Father
11 13 14 16 1 2 3 4 5 6
Family
ZFP35
6.8 6.9 7.0 7.1 7.2 7.3
ATP5SL
7.6 7.8 8.0 8.2

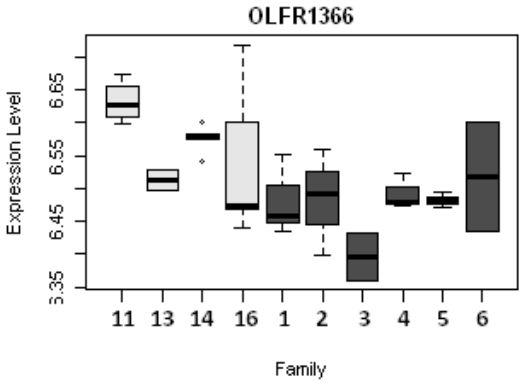

TSPAN3
Expression Level
9.4
9.6
9.8
10.0
Young
Old
Group

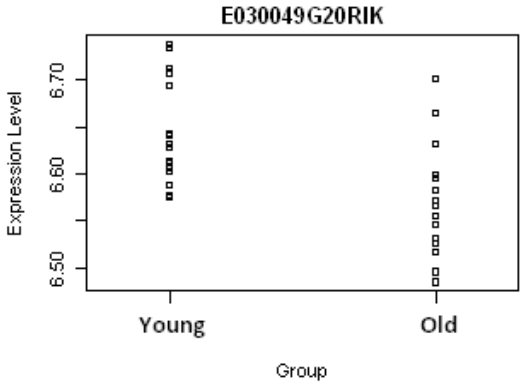

TSPAN3
Expression Level
9.4
9.6
9.8
10.0
Young Father
Old Father
11 13 14 16 1 2 3 4 5 6
Family

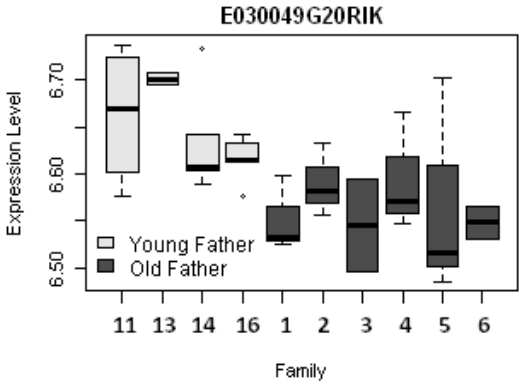

OLFR92
Expression Level
6.45
6.55
6.65
Young
Old
Group


OLFR92
Expression Level
6.45
6.55
6.65
11 13 14 16 1 2 3 4 5 6
Family


OLFR1366
Expression Level
6.35
6.45
6.55
6.65
Young
Old
Group


OLFR1366
Expression Level
6.35
6.45
6.55
6.65
11 13 14 16 1 2 3 4 5 6
Family


E030049G20RIK
Expression Level
6.50
6.60
6.70
Young
Old
Group


E030049G20RIK
Expression Level
6.50
6.60
6.70
Young Father
Old Father
11 13 14 16 1 2 3 4 5 6
Family

6.5.1.2 Post-Batch Correction

Interestingly, after array batch-effects had been corrected for, 876 genes (notably higher than in the analysis of pre-batch corrected samples) reached an uncorrected p-value of less than 0.05. The genes demonstrating a p-value below 0.001 from analysis between the offspring of young fathers and offspring of old fathers are shown in *Table 6.4* and the top ten differentially expressed genes shown in *Figure 6.9*. MUC15 was the top-ranked gene (p = 7.2E-06, q = 0.13). Of note, the rank order of genes was not largely affected by batch correction.

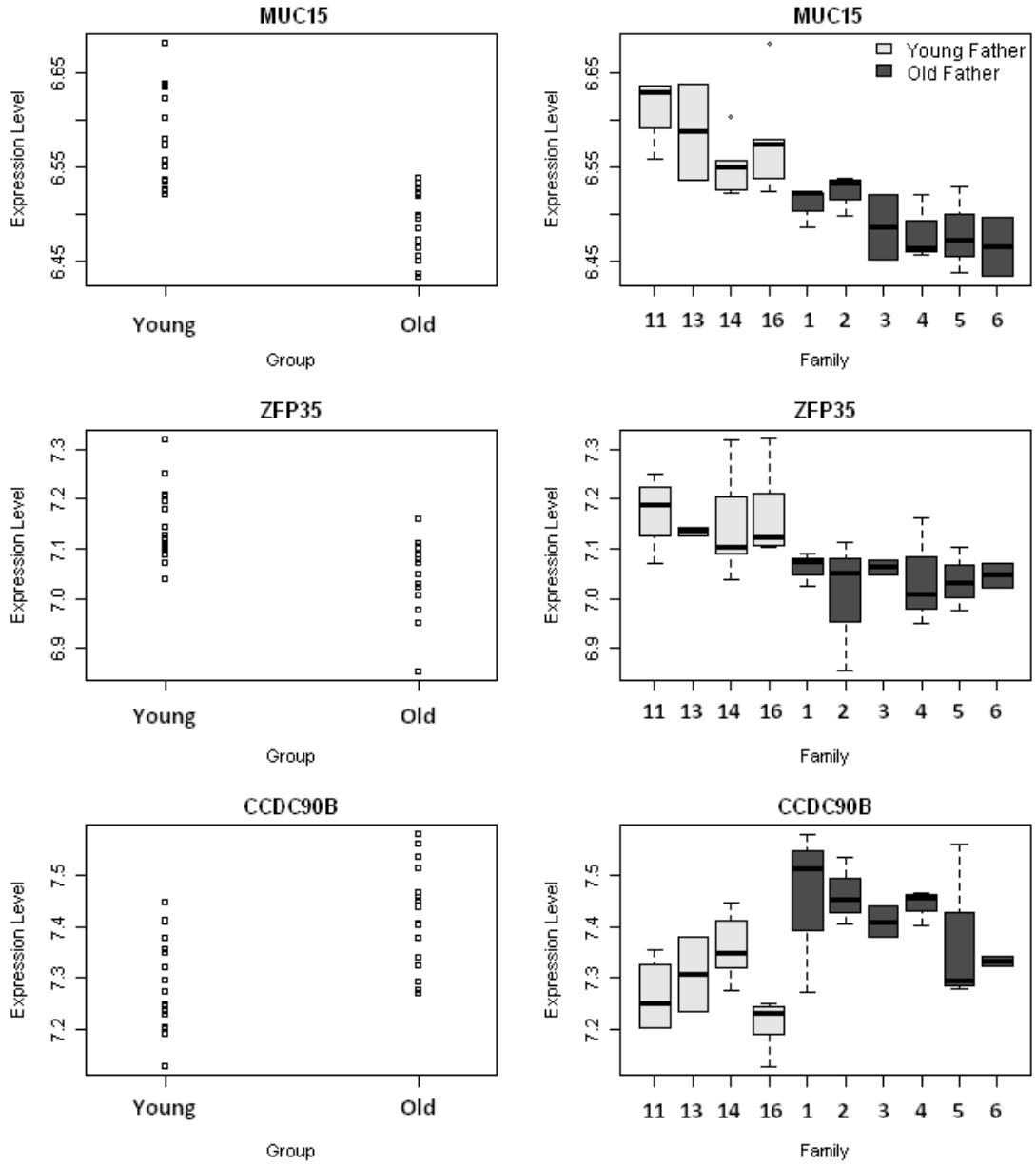
Table 6.4 - Differentially Expressed Genes from lumi Package - Post-Batch Correction

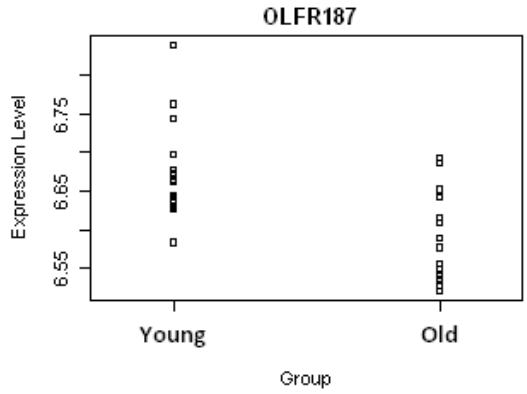
Genes which showed an uncorrected p-value of ≤ 0.001. The rank position from the pre-batch corrected data is included.

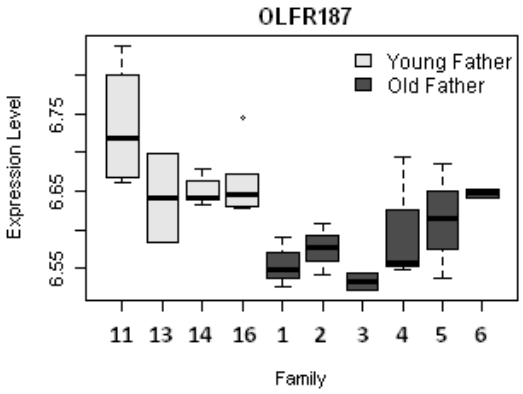
Gene	Pre-batch Corrected rank	Average	Average Young	Average Old	t	Uncorrected p-Value
MUC15	1	6.54	6.58	6.49	5.28	7.2E-06
ZFP35	5	7.10	7.16	7.04	4.46	8.4E-05
CCDC90B	2	7.35	7.28	7.42	-4.24	1.6E-04
OLFR187	4	6.63	6.67	6.59	4.10	2.4E-04
OLFR92	8	6.58	6.54	6.61	-4.06	2.7E-04
E130303B06RIK	3	6.73	6.77	6.68	4.05	2.8E-04
ZC3H18	56	7.22	7.17	7.26	-3.92	4.0E-04
ATP5SL	6	7.92	8.00	7.84	3.91	4.1E-04
TNFSF11	14	6.61	6.57	6.65	-3.77	6.1E-04
TSPAN3	7	9.78	9.71	9.86	-3.76	6.3E-04
EAR12	17	6.56	6.60	6.53	3.71	7.2E-04
LOC100046861	20	6.63	6.66	6.59	3.70	7.5E-04
RP9	30	8.48	8.53	8.43	3.69	7.6E-04
VTCN1	21	6.62	6.66	6.59	3.66	8.4E-04
OLFR1366	9	6.52	6.57	6.48	3.64	8.8E-04
FTL2	19	6.59	6.56	6.61	-3.63	9.2E-04
NSFL1C	31	7.85	7.93	7.78	3.61	9.6E-04
LARP4	12	6.55	6.58	6.53	3.58	0.001
CCL26	34	6.53	6.57	6.49	3.55	0.001
IER5L	13	7.77	7.84	7.70	3.55	0.001
C330005M16RIK	25	6.66	6.70	6.62	3.50	0.001
CD177	15	6.57	6.59	6.54	3.50	0.001
XKR8	54	6.98	6.94	7.03	-3.47	0.001

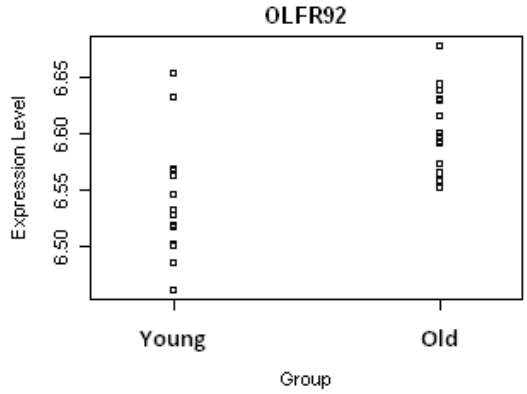
Figure 6.9 - Boxplot of Top 10 Differentially Expressed Genes Post-Batch Correction

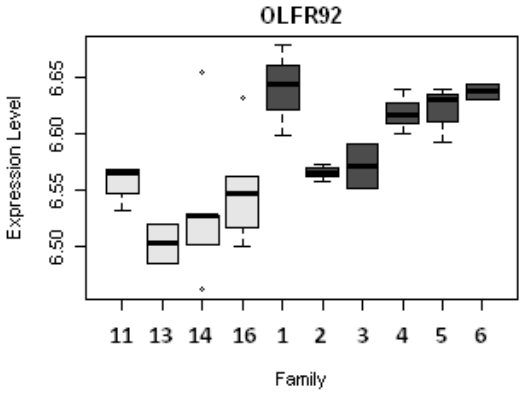
Top 10 differentially expressed genes from lumi analysis after array batch correction. All genes show strip charts of offspring of young fathers and offspring of old fathers as well as box plot of expression level by family. Black line in box plots shows median value, edges of boxes show upper and lower quartiles and whiskers show the upper and lower extreme values.

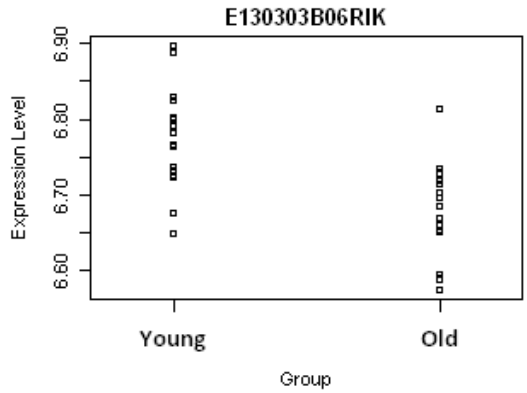


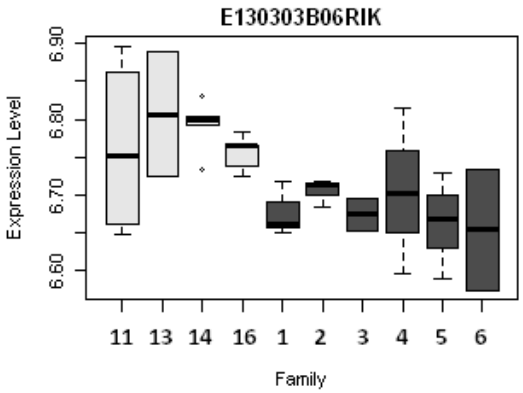

OLFR187
Expression Level
6.55
6.65
6.75
Young
Old
Group

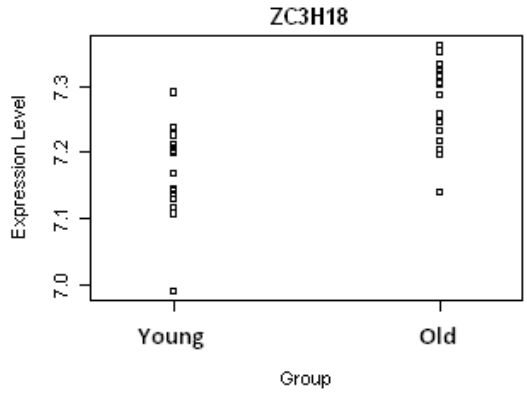

OLFR187
Young Father
Old Father
Expression Level
6.55
6.65
6.75
11 13 14 16 1 2 3 4 5 6
Family

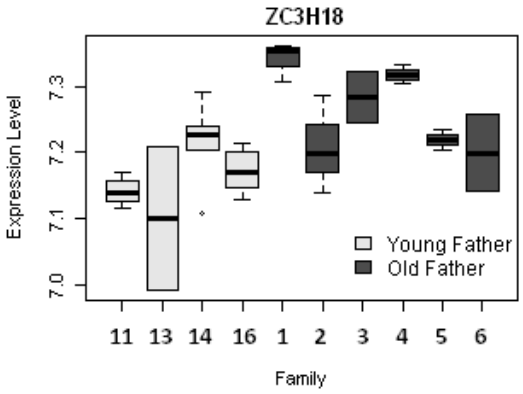

OLFR92
Expression Level
6.50
6.55
6.60
6.65
Young
Old
Group

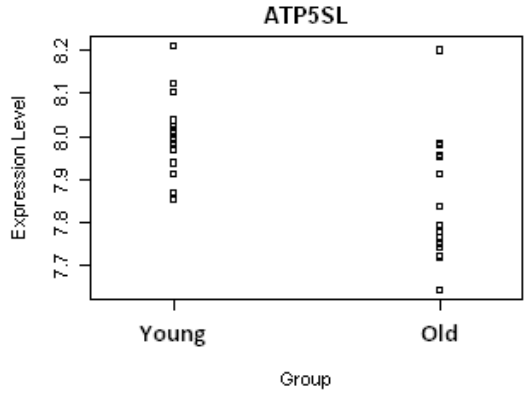

OLFR92
Expression Level
6.50
6.55
6.60
6.65
11 13 14 16 1 2 3 4 5 6
Family

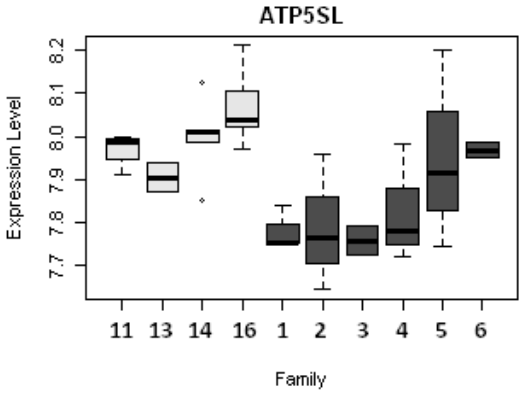

E130303B06RIK
Expression Level
6.60
6.70
6.80
6.90
Young
Old
Group

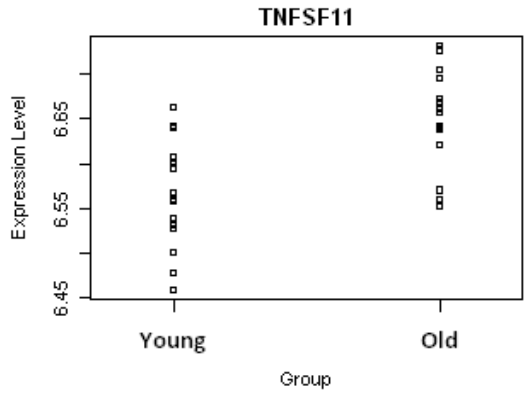

E130303B06RIK
Expression Level
6.60
6.70
6.80
6.90
11 13 14 16 1 2 3 4 5 6
Family

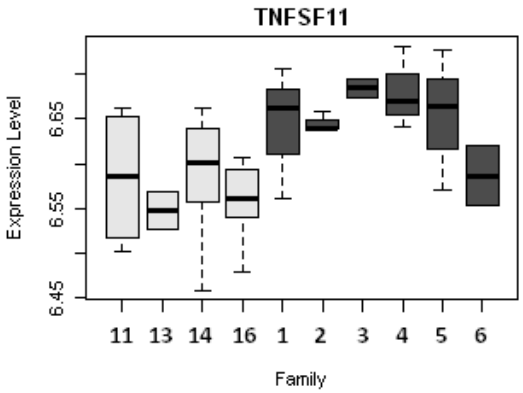

ZC3H18
Expression Level
7.0
7.1
7.2
7.3
Young
Old
Group

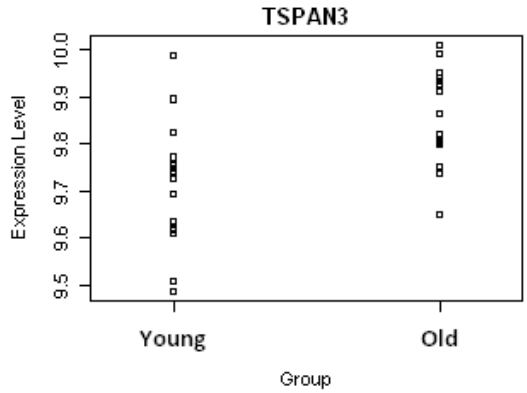

ZC3H18
Expression Level
7.0
7.1
7.2
7.3
Young Father
Old Father
11 13 14 16 1 2 3 4 5 6
Family

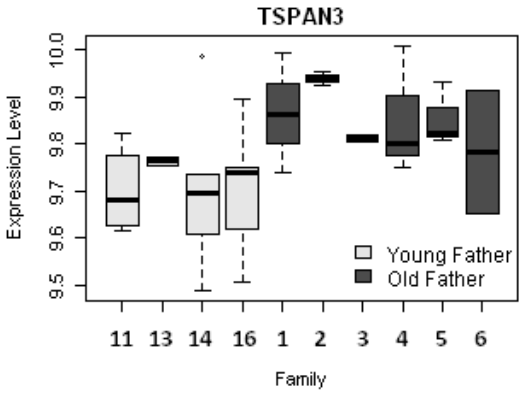

ATP5SL
Expression Level
7.7
7.8
7.9
8.0
8.1
8.2
Young
Old
Group


ATP5SL
Expression Level
7.7
7.8
7.9
8.0
8.1
8.2
11 13 14 16 1 2 3 4 5 6
Family


TNFSF11
Expression Level
6.45
6.55
6.65
Young
Old
Group


TNFSF11
Expression Level
6.45
6.55
6.65
11 13 14 16 1 2 3 4 5 6
Family

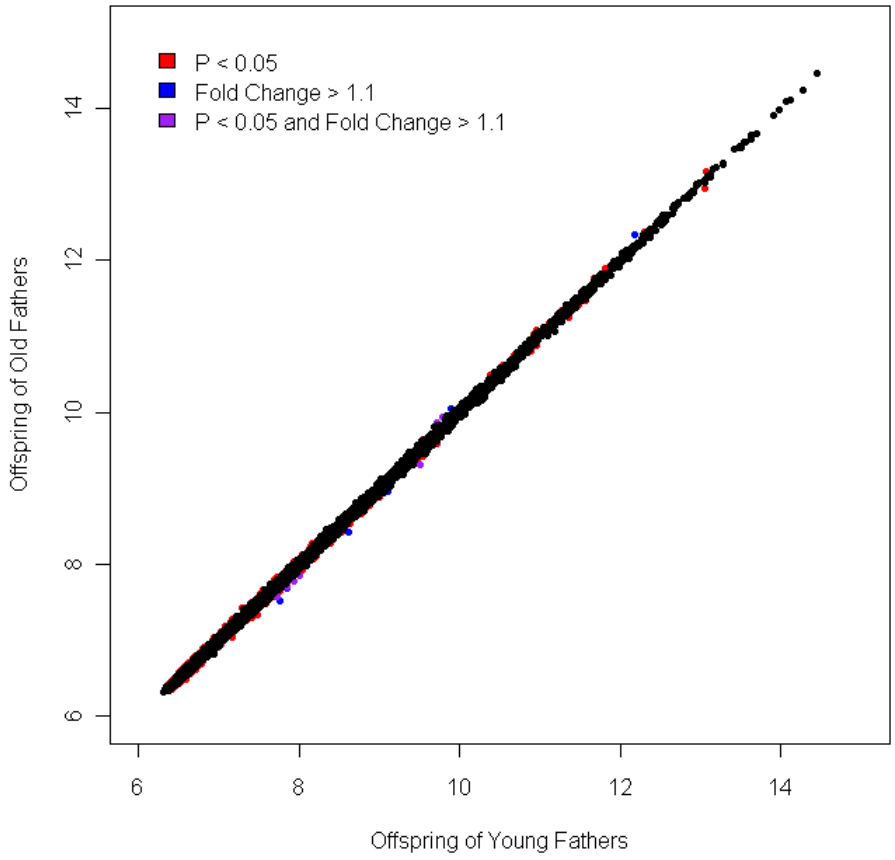

TSPAN3
Expression Level
9.5
9.6
9.7
9.8
9.9
10.0
Young
Old
Group


TSPAN3
Expression Level
9.5
9.6
9.7
9.8
9.9
10.0
Young Father
Old Father
11 13 14 16 1 2 3 4 5 6
Family

Overall, frontal cortex gene expression profiles were highly correlated (r^2 = 0.9997, p < 2.2E-16) between groups (*Figure 6.10*), suggesting that there are no systemic biological (or technical) differences between them as expected, and that differential expression is restricted to specific loci.

Figure 6.10 – Log2 Intensities of Offspring of Young Fathers by Offspring of Old Fathers

Average log2 intensity values of offspring of old fathers against offspring of young fathers. These genes were analysed with Ingenuity Pathway Analysis (section 6.5.5).



Log2 transformed fold-changes were transformed back into actual fold change and ranked. No genes reached a fold change of 1.5, which is the cut off for many disease-based gene expression studies, suggesting that the differences in gene expression between the offspring of young and old fathers are subtle. Genes reaching a fold change of over 1.1 are shown in *Table 6.5* along with their p-values. The largest fold change between the offspring of old and young fathers was S100A8 which had an 18% reduction in expression in the offspring

of old fathers compared to the offspring of young fathers. However, this gene did not have an uncorrected p-value of less than 0.05 so was not included in pathway analysis. The highest fold change which also had an uncorrected p-value of less than 0.05 was YWHAG with a reduction in expression of 14% in the offspring of old fathers.

Table 6.5 - Genes with a Fold Change over 1.1

Fold changes were calculated from log2 fold changes by calculating 2^x where x was the log2 fold change. P-values were calculated using lumi.

All Genes with Fold Change ≥ 1.1

Gene	p-value	Fold Change
S100A8	0.25	-1.18
TMEM10	0.07	-1.14
YWHAG	0.04	-1.14
RGS9	0.06	-1.12
RNF166	0.01	-1.12
ATP5SL	4.08E-04	-1.12
BCAS1	0.12	1.11
NSFL1C	9.61E-04	-1.11
TSPAN3	6.29E-04	1.11
LYPD1	0.07	-1.11
DNAJC15	0.05	1.11
IER5L	1.14E-03	-1.10
GNG10	0.07	1.10
CCDC90B	1.56E-04	1.10

All Genes with Fold Change ≥ 1.1 and p <0.05

Gene	p-value	Fold Change
YWHAG	0.04	-1.14
RNF166	0.01	-1.12
ATP5SL	4.08E-04	-1.12
NSFL1C	9.61E-04	-1.11
TSPAN3	6.29E-04	1.11
DNAJC15	0.05	1.11
IER5L	1.14E-03	-1.10
CCDC90B	1.56E-04	1.10

6.5.2 Family Based Analysis

From plotting the top differentially expressed genes by p-value (*Figure 6.9*), the expression differences seen in the offspring of old vs. young families appear to be variable by family, suggesting that family-specific changes may be seen in the context of advanced paternal age. Due to this, the offspring of old fathers were split by family and compared to the offspring of young fathers. Although several genes met the threshold set in section 6.5.1.2, different families had differentially expressed genes. Results of this analysis are shown in *Table 6.6*.

Table 6.6 – Differentially Expressed Genes by Family in Offspring of Old Fathers

Differentially expressed genes in offspring of old father families compared to offspring of young fathers with p-value below 0.001 and fold change over 1.1. Rank indicates rank by p-value in post-batch corrected analysis.

Family 1

Rank	Gene	Fold Change	p-value
20	IER5L	1.21	8.97E-05
3328	TMEM183A	1.13	9.42E-05
5004	PAFAH1B3	1.14	1.06E-04
1216	SEC15L1	1.13	2.00E-04
7309	PIK3R2	1.10	2.26E-04
606	GSK3A	1.10	2.95E-04
11982	ACTR1B	1.14	3.44E-04
6046	LOC666676	1.23	3.77E-04
100	NFU1	1.13	4.24E-04
7	ZC3H18	1.13	6.07E-04
436	OGFOD2	1.10	6.11E-04
99	LOC100046163	1.11	7.62E-04
8	ATP5SL	1.17	8.73E-04
242	EIF3G	1.19	0.001
1404	RAD23B	1.33	0.001
2576	SLITRK2	1.11	0.001
2141	COQ7	1.14	0.001
435	STT3B	1.18	0.001
707	HERPUD1	1.17	0.001

Family 2

Rank	Gene	Fold Change	p-value
4199	GEM	1.12	8.22E-05
36	QPCTL	1.10	1.54E-04
8595	RAB34	1.10	6.54E-04
1747	RAB11FIP5	1.11	7.28E-04
1509	HRASLS	1.18	7.44E-04
2157	ZFP825	1.10	7.80E-04
13	RP9	1.13	9.50E-04
927	4921530G04RIK	1.12	0.001
3692	THEM4	1.14	0.001

Family 3

Rank	Gene	Fold Change	p-value
28	PSCD4	1.16	3.67E-05
3241	KIF11	1.14	9.80E-05
1416	PDE7B	1.15	2.06E-04
15346	ZFP759	1.11	2.69E-04
6053	EG435391	1.12	2.84E-04
174	TM4SF5	1.14	3.43E-04
144	AMELX	1.14	3.46E-04
1698	2010110P09RIK	1.12	3.66E-04
1750	TAC4	1.11	4.20E-04
2432	ASCL1	1.14	4.81E-04
7962	PAPLN	1.11	5.01E-04
2438	1700019L03RIK	1.11	5.40E-04
1503	SLC27A2	1.15	6.59E-04

Family 4

Rank	Gene	Fold Change	p-value
1027	TALDO1	1.13	1.66E-04
107	RSPO4	1.12	3.72E-04
8518	HMG20B	1.11	7.77E-04
6820	WIF1	1.10	8.93E-04
170	MTIF2	1.11	0.001
109	SLC16A11	1.12	0.001

Family 5

Rank	Gene	Fold Change	p-value
61	UGT1A10	1.14	2.79E-05
131	EPHA5	1.13	1.57E-04
8557	PDRG1	1.18	1.75E-04
11778	ANP32A	1.11	3.94E-04
15669	UTP6	1.10	8.48E-04
5775	FRMPD4	1.12	9.38E-04
11982	ACTR1B	1.13	9.56E-04
2481	CCDC109A	1.11	0.001
16961	VTI1B	1.13	0.001
15019	PPP1R16A	1.14	0.001
2267	TPD52L1	1.10	0.001

Family 6

Rank	Gene	Fold Change	p-value

1685	DUS1L	1.12	6.83E-04
13416	OLFR1298	1.10	6.99E-04
2231	LYZL4	1.14	7.82E-04
1754	PUNC	1.13	9.42E-04
259	ZFP459	1.10	0.001
129	RASGRF2	1.15	0.001
1623	HAND2	1.10	0.001
2610	4430402I18RIK	1.10	0.001
10496	LY6G6C	1.10	0.001
258	FCHSD2	1.12	0.001

469	OLFR412	1.15	2.90E-05
1698	2010110P09RIK	1.12	1.35E-04
5969	ETV6	1.11	2.29E-04
4087	GAN	1.15	0.001
17927	V1RF3	1.10	0.001

6.5.3 Relating Expression and Methylation

Given the link between DNA methylation and gene expression where methylation is negatively correlated with expression throughout the genome (Bell, Pai et al. 2011) (section 1.6), I was able to look specifically at the expression for the imprinted genes regulated by the DMRs assessed in Chapter 5. There were no differences between the offspring of young fathers and offspring of old fathers in Mcts2 DMRs in the frontal cortex (*Figure 5.23*), and when we look at the expression levels in this gene we see no significant differences between groups ($t = -0.84$, d.f. = 30, $p = 0.41$). Kcnq1ot showed trends towards significant DNA methylation differences between the offspring of young and old fathers in the frontal cortex and although it is non-coding it also controls expression of *CDKN1C*, however expression of *CDKN1C* was not different between groups ($t = -0.22$, d.f. = 30, $p = 0.83$). Gnas Nesp controls expression of *Nesp55*, but this was not covered on the Illumina Mouse Ref8 V2 BeadChips.

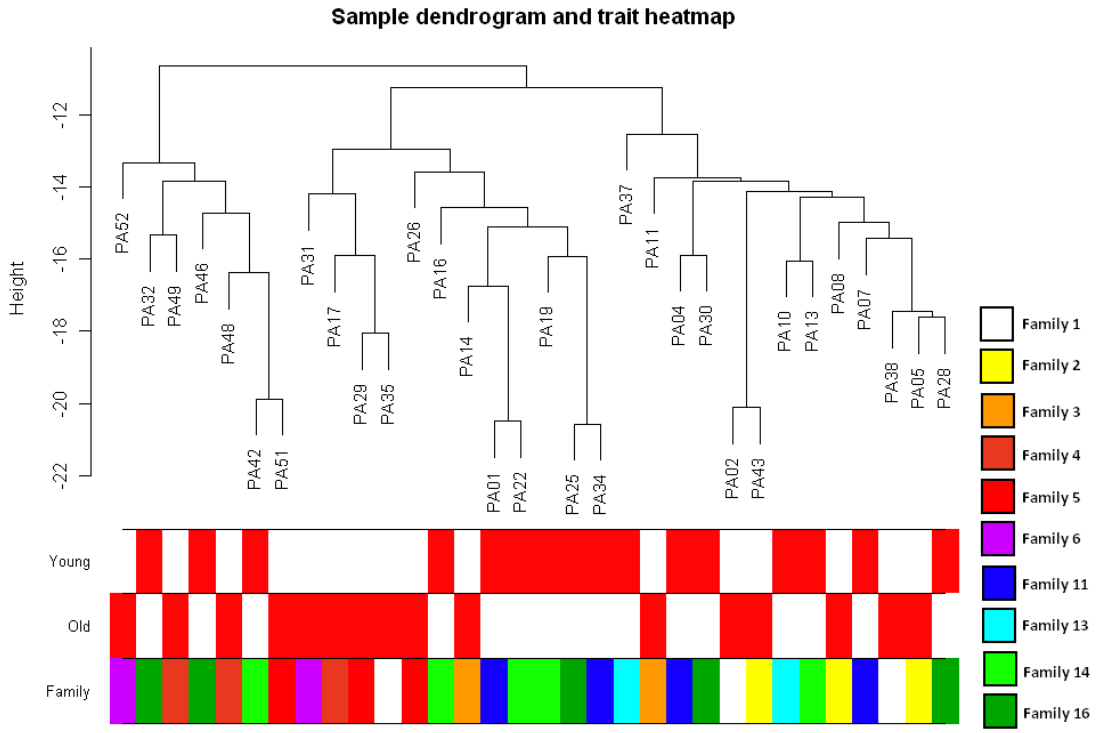
6.5.4 WGCNA

6.5.4.1 Sample Relationships

Sample relationships based on expression values are shown in *Figure 6.11* and are coloured according to group (offspring of young or old father) and family membership. Overall, the clustering of samples does not clearly distinguish between groups of offspring as a function of paternal age, concurring with the lumi results presented above and confirming that systemic transcriptomic differences are unlikely to distinguish between groups. There are, however, several clusters of samples grouped by high and low paternal age. Samples did not appear to cluster by family.

Figure 6.11 - Clustering Dendrogram of Samples

Clustering is based on their euclidean distance and was carried out using Ward's method which infers that the distance between two clusters is how much the sum of squares will increase when we merge them. The heat map at the bottom identifies which samples belong to certain groups i.e. offspring of old or young father and family.



6.5.4.2 Identifying Modules

WGCNA demonstrated that frontal cortex gene expression patterns were clearly modular, as shown previously in other gene expression datasets (Voineagu, Wang et al. 2011). Genes were assigned into 22 co-expression modules, with 8523 loci remaining unassigned to a module (*Table 6.7* and *Figure 6.12*). Module sizes ranged from 118 to 1146 genes.

Table 6.7 – Results of Network Construction in WGCNA

Table showing the number of genes per module based on WGCNA analysis. Genes were split into 22 modules with 8523 remaining unassigned to a module (module 0)

Module Number	0	1	2	3	4	5
Number of Members	8523	1146	1094	1013	972	745
Colour	grey	turquoise	blue	brown	yellow	green

Module Number	6	7	8	9	10	11
Number of Members	586	566	501	275	270	257
Colour	red	black	pink	magenta	purple	greenyellow

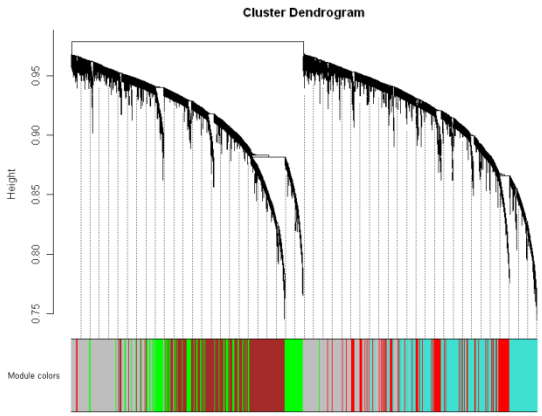
Module Number	12	13	14	15	16	17
Number of Members	256	242	230	202	200	199
Colour	tan	salmon	cyan	midnightblue	lightcyan	grey60

Module Number	18	19	20	21	22
Number of Members	197	191	159	155	118
Colour	lightgreen	lightyellow	royalblue	darkred	darkgreen

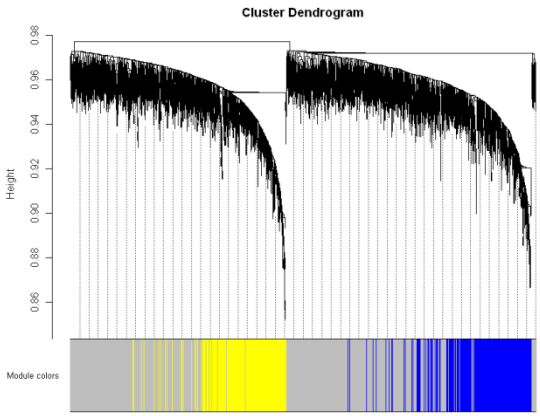
Figure 6.12 - Clustering Dendrogram of Genes

Dissimilarity based on topological overlap, together with assigned module colours. Data is divided into 22 modules (indicated by different colour) with 8523 remaining unassigned to a module (light grey)

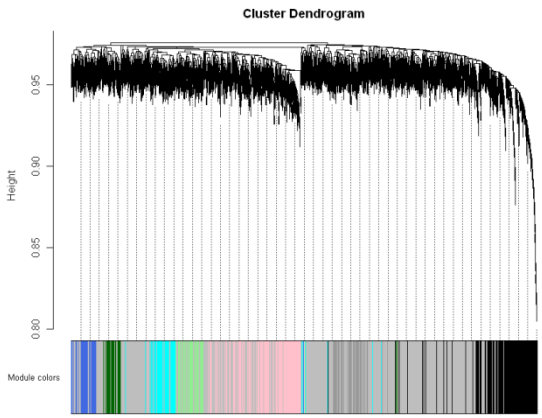
Block 1



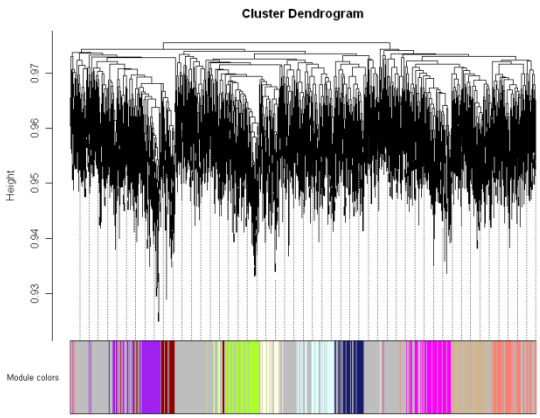
Block 2



Block 3



Block 4



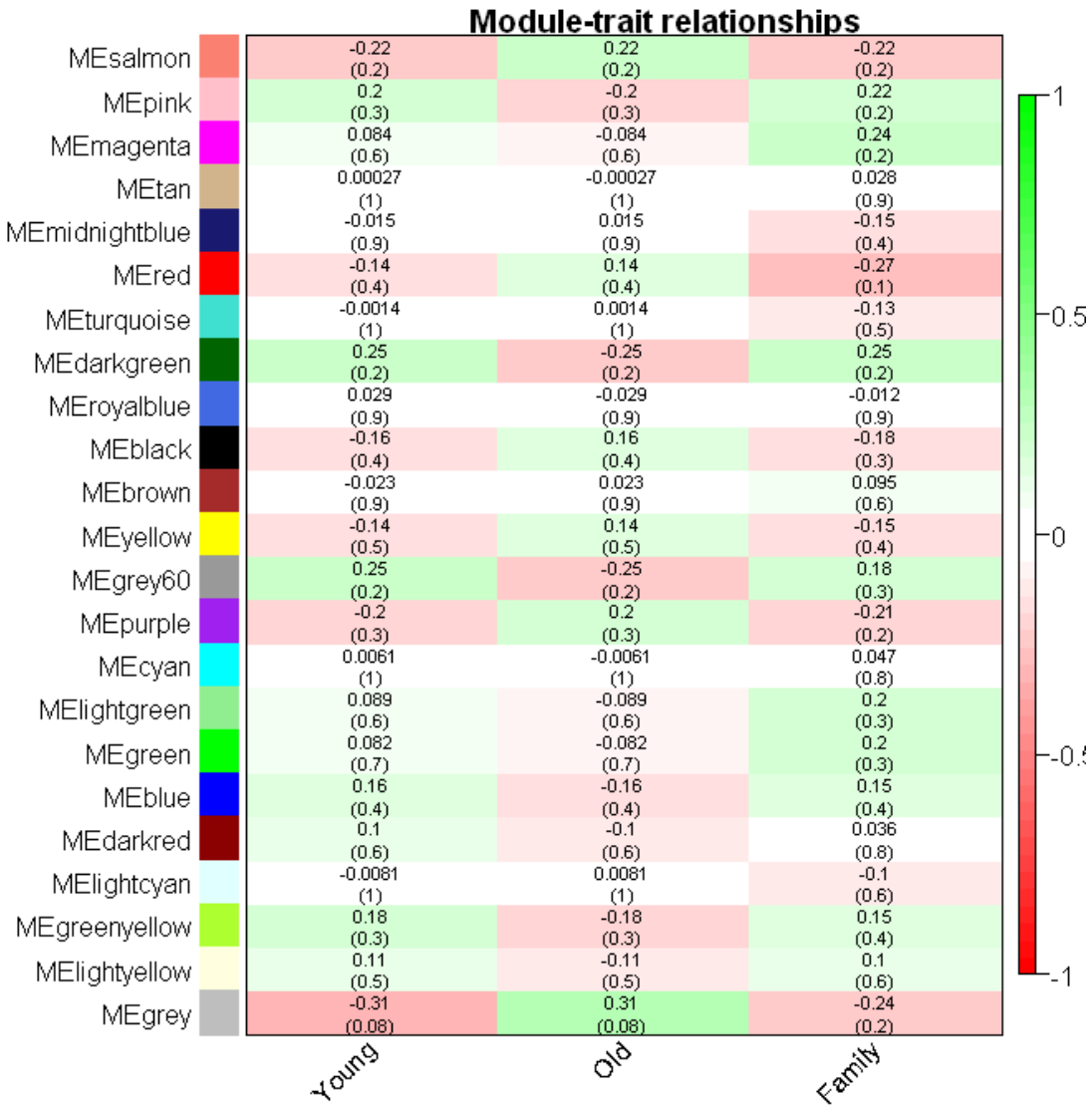
6.5.4.3 Relating Modules

Using WGCNA analysis and looking at module trait relationships (*Figure 6.13*), modules that have significant correlations with traits of the data can be identified. After tabling the correlations and respective p-values to paternal age group and family traits, however, no module reaches a significant correlation with either. This suggests that if paternal age is

having an effect on gene expression in the frontal cortex, then this appears to be specific to certain loci and not altering transcription across entire transcriptional modules of co-expressed genes.

Figure 6.13 – Module Trait Relationships

Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value in brackets. The table is colour-coded by correlation of the module to a trait according to the colour legend on the green with red being a positive correlation and red being a negative correlation.



6.5.5 Pathway Analysis

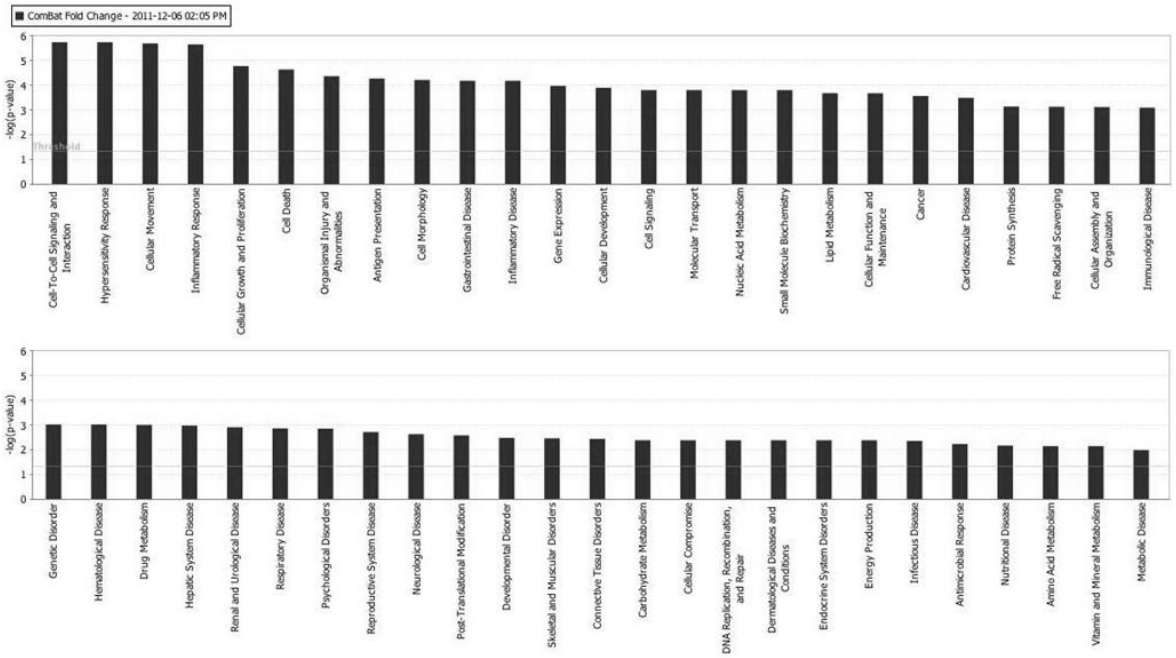
Pathway analysis was carried out using Ingenuity Pathway Analysis (IPA), a powerful package that identifies relationships, mechanisms, functions, and pathways in experimental data. Pathway analysis using IPA was carried out on genes which reached an uncorrected p-value less than 0.05 from post batch corrected lumi analysis. The number of genes which were below this value was 876, of which 605 were successfully mapped to annotated loci within the Ingenuity Knowledge database. 25 pathways were identified from the genes put into pathway analyses (*Table 6.8* and *Figure 6.14*) and the top five pathways are shown in *Figure 6.15*. In a previous studies of gene expression and aging, an up-regulation of inflammatory response genes was observed (Lee, Weindruch et al. 2000), and we also observe differential expression of a gene network related to inflammatory response. Genes involved in the immune system appear to be over represented in those which show an expression change with increased paternal age with 45 different genes involved in inflammatory response. Other pathways highlighted by IPA are DNA replication, recombination and repair, of which gene expression changes in genes associated with DNA repair has been previously associated with aging in mouse brain (Lee, Weindruch et al. 2000) and spermatogonial cells (Kokkinaki, Lee et al. 2010).

Table 6.8 – Networks Identified from IPA of Most Differentially Expressed Genes

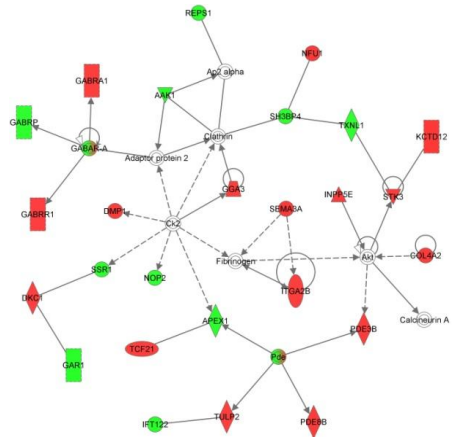
Score	Focus Molecules	Top Functions
38	26	Cellular Assembly and Organization, Cellular Function and Maintenance, Organismal Injury and Abnormalities
38	26	Cell Death, Cell Morphology, Cellular Function and Maintenance
34	24	Inflammatory Response, Cellular Movement, Immunological Disease
30	23	Gene Expression, Cell Cycle, Decreased Levels of Albumin
30	22	Post-Translational Modification, Cellular Development, Hematopoiesis
29	22	Gene Expression, Protein Synthesis, Carbohydrate Metabolism
27	21	Cell Signalling, Molecular Transport, Nucleic Acid Metabolism
24	20	Cellular Movement, Cell Signalling, Molecular Transport
22	16	Cellular Assembly and Organization, Cell-To-Cell Signalling and Interaction, Cellular Function and Maintenance
22	18	Cell-To-Cell Signalling and Interaction, Cellular Movement, Immune Cell Trafficking
21	17	Cell Cycle, DNA Replication, Recombination, and Repair, Cellular Assembly and Organization
20	19	DNA Replication, Recombination, and Repair, Cell-To-Cell Signalling and Interaction, Nervous System Development and Function
20	17	Cellular Development, Hematopoiesis, Cell-To-Cell Signalling and Interaction
20	17	Cell Death, Haematological System Development and Function, Cellular Development
19	16	Cell Signalling, Small Molecule Biochemistry, Immunological Disease
17	15	Cell Death, Lipid Metabolism, Small Molecule Biochemistry
16	14	Connective Tissue Development and Function, Skeletal and Muscular System Development and Function, Tissue Morphology
16	14	Cell Cycle, Cellular Development, Connective Tissue Development and Function
16	14	Nervous System Development and Function, Carbohydrate Metabolism, Lipid Metabolism
15	14	Cellular Development, Haematological System Development and Function, Hematopoiesis
15	14	Cellular Development, Connective Tissue Development and Function, Skeletal and Muscular System Development and Function
15	14	Inflammatory Response, Cell Morphology, Cellular Movement
15	14	Cancer, Gastrointestinal Disease, Lymphoid Tissue Structure and Development
14	13	Post-Translational Modification, Protein Degradation, Protein Synthesis
14	13	DNA Replication, Recombination, and Repair, Energy Production, Nucleic Acid Metabolism

Figure 6.14 – Networks of Most Differentially Expressed Genes

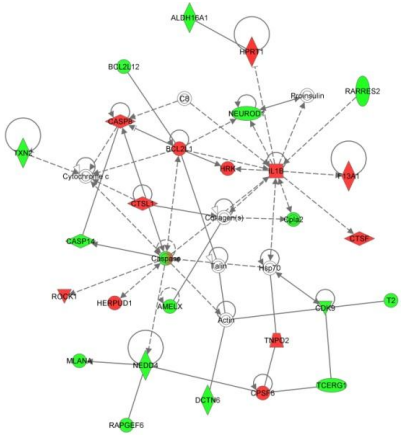
IPA analysis of most differentially expressed genes with fold change over 1.1



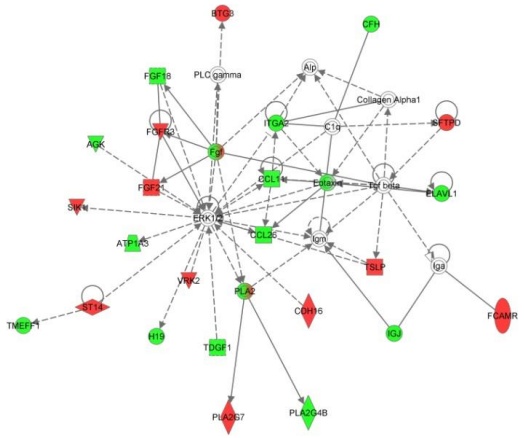
1. Cellular Assembly and Organization, Cellular Function and Maintenance Organismal Injury and Abnormalities



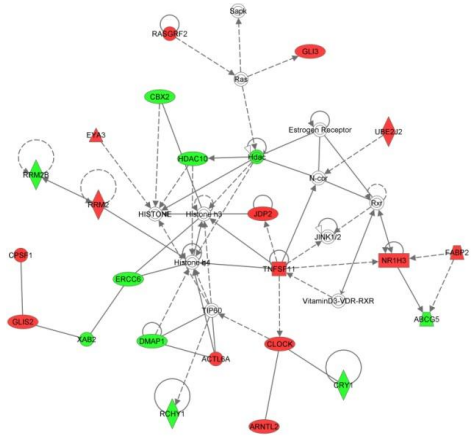
2. Cell Death, Cell Morphology, Cellular Function and Maintenance



3. Inflammatory Response, Cellular Movement, Immunological Disease



4. Gene Expression, Cell Cycle, Decreased Levels of Albumen



5. Post-Translational Modification, Cellular Movement, Hematopoiesis

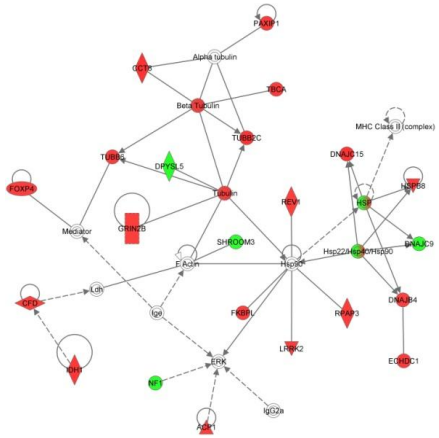


Figure 6.15 – Top Five Networks from IPA

Green indicates genes which were significantly down-regulated in the offspring of old males in this study and red indicated genes which were significantly up-regulated in the offspring of old fathers in this study.

6.5.6 Relating CNVs to Expression

As CNVs have been shown to affect expression of the genes they encompass (Stranger, Forrest et al. 2007) (section 4.2.1), expression levels of the genes affected by CNVs identified in Chapter 4 were investigated. To test the hypothesis that CNVs may directly influence levels of gene expression, the initial analysis focused on CNVs observed that had been previously been reported in other mouse studies as these are more likely to be truly positive CNVs. One CNV from Watkins-Chow *et al's* study (Watkins-Chow and Pavan 2008), was seen in six offspring, for whom we also had frontal cortex gene expression data (section 4.5.1). All other 'common' CNVs (i.e. those reported in other studies (see *Figure 4.21*, *Figure 4.22* and *Figure 4.23*) were either i) detected in animals for whom gene expression data was not collected, ii) affected genes not contained on the Illumina Mouse Ref8 V2 BeadChips, or iii) were found to be not expressed in the frontal cortex. This duplication contains the gene *Ide,* which is covered by the Illumina Mouse Ref8 V2 BeadChips and expressed at detectable levels in the frontal cortex tissue examined. When comparing the log2 values for *Ide* between CNV carriers (n=6) and non-carriers (n=26), a significant difference (t = 4.75, d.f. = 31, p = 0.002) was seen, with duplication carriers having higher expression, with a fold change of 1.26 (*Figure 6.16*).

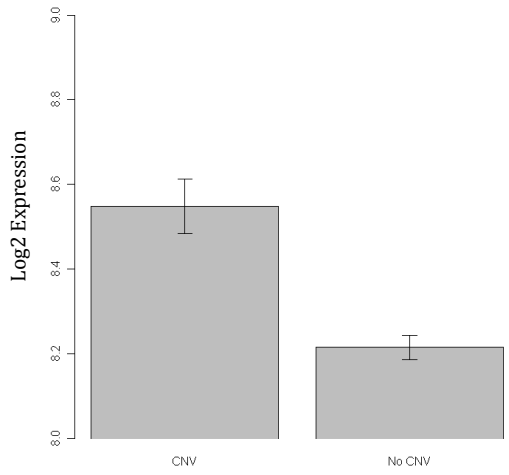


Figure 6.16 – Expression of Ide by CNV Status

Log2 expression between carriers and non-carriers of a CNV spanning Ide on chromosome 19 with bars representing SEM. Fold change equates to 1.26 in the group containing the CNV.

Next, genes affected by other CNVs detected in our array-CGH experiments were investigated which were present in 1-3 offspring individuals. The total number of genes affected by CNVs for which gene expression data was available was 179 across all offspring (section 4.4.8). Although CNVs have been shown to affect the expression of genes up to 0.5Mb away, genes in the direct vicinity of a genomic alteration are likely to show a greater gene expression change (Henrichsen, Vinckenbosch et al. 2009) and because of this, the analysis was restricted only to genes directly encompassed by the CNV. Strikingly, 133 genes out of 179 genes (74%) showed a significant gene expression difference between CNV carriers and non-carriers. The genes with a fold change in expression over 1.1 related to CNVs can be seen in *Table 6.9*. Expression, p-values and fold changes for all other genes are shown in Appendix 4. Barplots of the top 10 genes differentially-expressed CNV-associated genes are shown in *Figure 6.17*. Interestingly, although all observed deletions are associated with a decrease in expression, not all Duplications show an increase in expression indicating that the relationship between gene dosage and gene expression is not always straightforward.

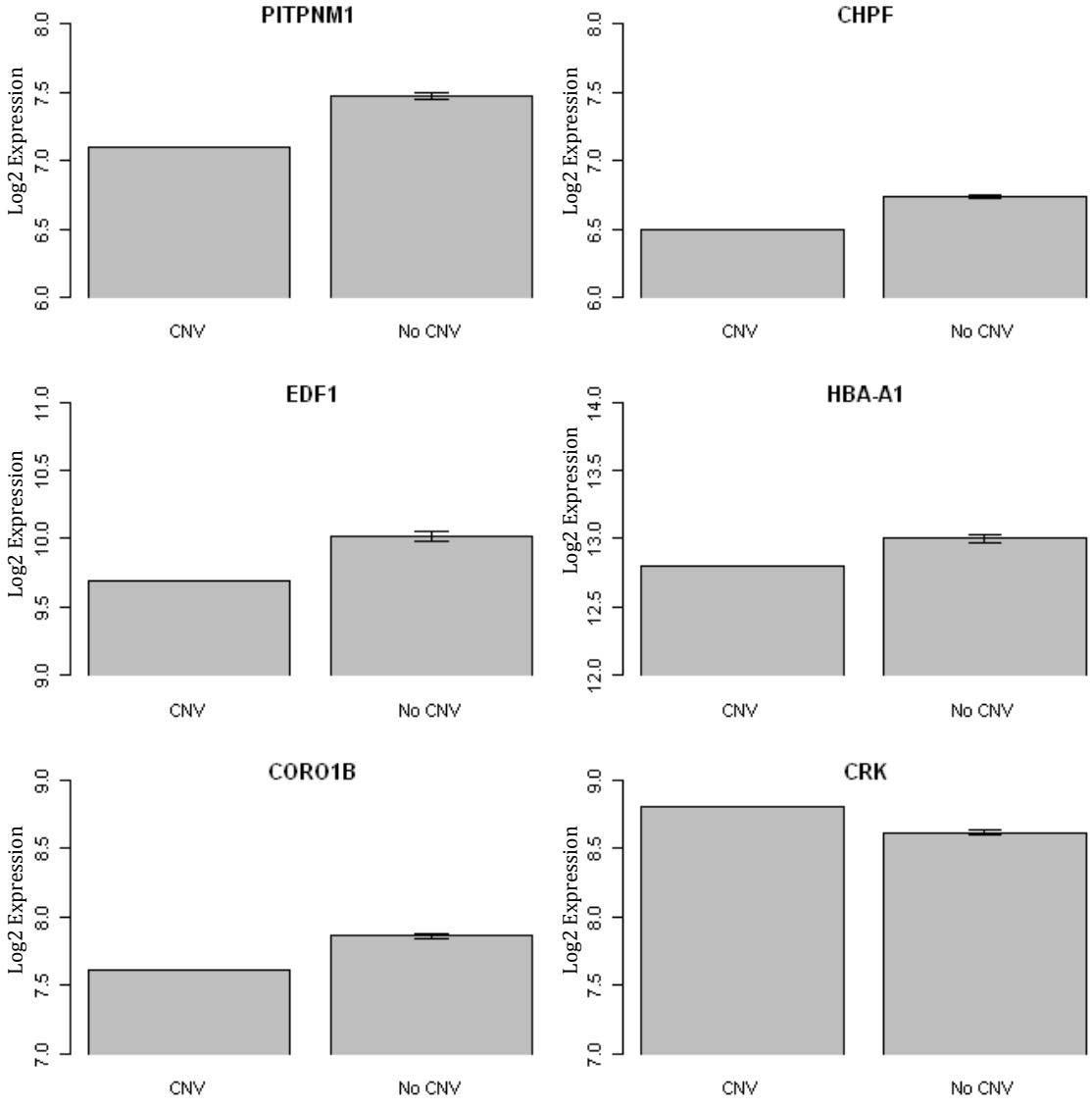
Table 6.9 – Relating CNVs to Expression

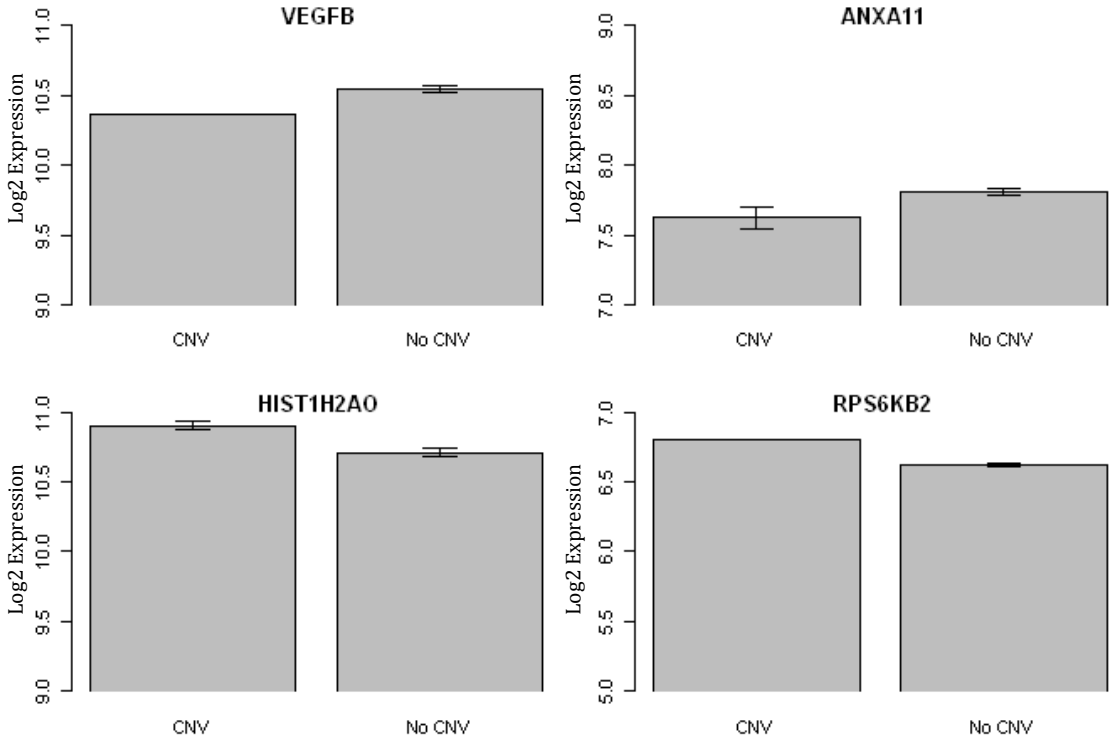
Table showing log2 expression values between individuals containing CNVs in the listed genes and those without for genes which show a fold change ≥ 1.1. 'Average' is the average expression across all individuals regardless of CNV status.

Gene	Number of Individuals with CNV	Average CNV	Average	Fold Change	t-value	p-value	Direction of Change	CNV Type
PITPNM1	1	7.10	7.46	1.28	6.73	1.57E-07	↓	Duplication
EDF1	1	9.69	10.00	1.24	-2.10	0.044	↓	Duplication
IFIT3	1	7.89	7.59	1.23	1.93	0.063	↑	Duplication
CORO1B	1	7.61	7.85	1.18	-8.10	3.82E-09	↓	Duplication
CHPF	1	6.49	6.73	1.18	4.85	3.26E-05	↓	Duplication
D14ERTD449E	2	7.33	7.56	1.17	1.94	0.061	↓	Deletion
HBA-A1	1	12.80	13.00	1.15	-6.08	9.74E-07	↓	Deletion
CRK	1	8.81	8.62	1.14	3.97	3.93E-04	↑	Duplication
VEGFB	1	10.36	10.54	1.13	-4.27	1.73E-04	↓	Duplication
HIST1H2AO	3	10.90	10.72	1.13	-4.01	3.60E-04	↑	Duplication
ANXA11	2	7.63	7.80	1.13	-2.25	0.031	↓	Deletion
RPS6KB2	1	6.80	6.63	1.13	6.85	1.11E-07	↑	Duplication
TMEM198	1	8.63	8.46	1.12	-7.09	5.85E-08	↑	Duplication
TBC1D5	1	7.46	7.61	1.11	7.88	6.81E-09	↓	Duplication
PITPNM3	1	8.20	8.35	1.11	-4.37	1.29E-04	↓	Duplication
DBNDD1	1	7.34	7.19	1.11	7.37	2.66E-08	↑	Duplication
UBE3B	1	9.06	8.91	1.11	14.72	1.56E-15	↑	Duplication
EFNA2	1	6.65	6.50	1.11	0.00	1.000	↑	Duplication
CDK2AP2	1	8.12	7.98	1.10	0.05	0.963	↑	Duplication
MACROD1	1	7.26	7.40	1.10	2.72	0.011	↓	Duplication
IL4I1	1	6.65	6.78	1.10	-1.72	0.095	↓	Duplication
A630095E13RIK	1	6.47	6.60	1.10	5.35	7.86E-06	↓	Duplication

Figure 6.17 – Top 10 Fold Changes in Genes Affected by CNVs

Log2 expression split by CNV status of the listed gene for the top 10 genes which show a fold change over 1.1 and a p-value below 0.05. Error bars represent SEM.





6.6 Discussion

6.6.1 Summary of Results

From between group comparisons, 876 genes showed an uncorrected p-value of less than 0.05. However, after FDR correction only MUC15 showed a q value below 0.2 and none showed a q-value below 0.1. Therefore no gene reached high confidence in terms of gene expression between groups and no genes reached a fold change of 1.5 from the offspring of young fathers to the offspring of old fathers. The highest fold change between groups was 18% in S100A8; however, this gene did not have an uncorrected p-value of less than 0.05. The highest fold change which also had an uncorrected p-value of less than 0.05 was YWHAG with a reduction in expression of 14% in the offspring of old fathers and an uncorrected p-value of 0.04. Although gene expression in the frontal cortex was shown to be modular using WCNGA, no modules showed significant differences between paternal age groups.

Of particular interest given previous gene expression analyses of aging, pathway analysis on the list of nominally-significant genes revealed enrichment for gene pathways involved in

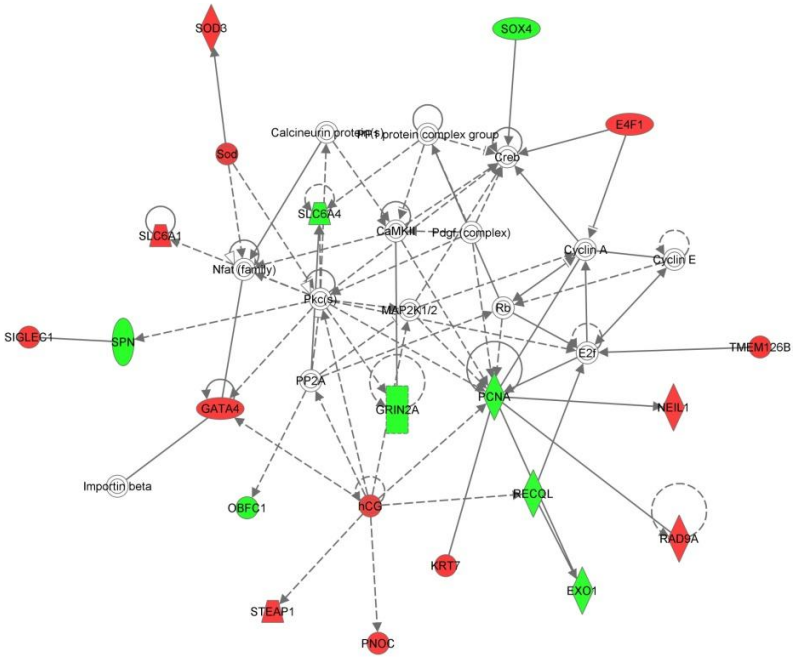
inflammatory and immunological processes. Also enriched were networks involved in cellular assembly and organization, cellular function and maintenance organismal injury and abnormalities, cell death, cell morphology, cellular function and maintenance, inflammatory response, cellular movement, immunological disease, gene expression, cell cycle, decreased levels of albumen and post-translational modification, cellular movement, hematopoiesis. Genes also involved in DNA replication, recombination, and repair were enriched. Finally, a large percentage of genes which show an expression change when contained within a CNV were observed.

6.6.2 Previous Literature

As mentioned in section 6.5.5, studies of aging in the brain have shown enrichment for genes associated with inflammatory response, and DNA repair (Lee, Weindruch et al. 2000) and studies on aging in spermatogonial cells also showed enrichment for changed in genes associated with DNA repair (Kokkinaki, Lee et al. 2010). Both these networks were also observed in our study. As differential expression of genes associated with DNA repair are also seen in spermatogonial cells, it may be that expression of these genes are affected in their offspring as well. Changes in expression in genes associated with DNA repair could mean that DNA contains more genomic alterations compared to the offspring of younger fathers. However, the genes in this network are a mixture of up and down-regulated genes (*Figure 6.18*) as opposed to all being in the same direction, and we have no information on small genomic alterations. After comparing genes which have altered gene expression with other studies, we observe some interesting overlap. Twelve genes identified in Paul *et al's* study of age related changed to gene expression in spermatocytes also showed altered gene expression between the offspring of old and young fathers (Paul, Nagano et al. 2011) and 83 genes overlapped with those showing differential expression in Kokkinani *et al's* study into spermatogonial cells from mice of different ages (Kokkinaki, Lee et al. 2010). There were no overlapping pathways between this study and Alter *et al's* study of expression changes associated with paternal age in the brain (Alter, Kharkar et al. 2011), but this could be due to this study taking place in RNA from frontal cortex and Alter *et al's* study being on RNA from hippocampus.

Figure 6.18 – IPA Network - DNA Replication, Recombination, and Repair, Cell-To-Cell Signalling and Interaction, Nervous System Development and Function

Green indicated genes which were down-regulated in the offspring of old males in this study and red indicated genes which were up-regulated in the offspring of old fathers in this study.



Gregg *et al's* case-control study on RNA from blood samples highlighted significant changes in genes associated with cell signalling, which mirrors the differences we observe in this study (Gregg, Lit et al. 2008). From the genes which reached an uncorrected p-value of less than 0.05, 31 genes were identified which have previously been associated with schizophrenia (*ATP1A3*, *C5orf13*, *CHRNA6*, *CHRNB1*, *CLOCK*, *DRD5*, *EBF1*, *FXYD6*, *GABRA1*, *GABRP*, *GPR37*, *GRIN2A*, *GRIN2B*, *HPRT1*, *HTR4*, *HTR7*, *IL10*, *IL1B*, *IL4*, *KCNK1*, *LTA*, *NDUFV1*, *PDE3B*, *PMP22*, *PNOC*, *RAPGEF6*, *SEMA3A*, *SLC6A1*, *SLC6A4*, *TUBB2C*, *TUBD1*) although these genes are normally only implicated in one study and so are not the most confident findings. Using the top genome-wide association results from genome-wide association studies for schizophrenia, none are significantly differentially expressed between the offspring of young and old fathers (Ripke, Sanders et al. 2011). Of genes which reached an uncorrected p-value of <0.001, only TNFSF11 has previously been implicated in the psychiatric disorders of interest with regards to paternal age. TNFSF11 is located at human locus 13q14 which has been identified in linkage studies as a shared locus of schizophrenia and BD (Maziade, Chagnon et al. 2009).

6.6.3 Most Differentially Expressed Genes and Functions

Looking at genes which showed a p-value below 0.05 and a fold change over 1.1 across all families, eight showed differences between the offspring of young and old male mice. *YWHAG* belongs to the family of YWHA (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation) proteins and mediates signal transduction through binding to phosphoserine-containing proteins. Genes in this family have been widely studied for their association with schizophrenia because of their role in neurotransmission (Berg, Holzmann et al. 2003). Deletion in *YWHAG* has previously been associated with developmental delay and intellectual disability (Ramocki, Bartnik et al. 2010) and with speech delay in Williams-Beuren syndrome (Berg, Brunetti-Pierri et al. 2007). *YWHAG* also has an association with longevity in humans (Deelen, Uh et al. 2011). No other genes in this list demonstrate convincing evidence for their role in aging or psychiatric conditions in the current literature.

The analysis of differentially expressed genes in individual families identified several neuropsychiatric candidate loci. This suggests that rare family-specific events (perhaps mediated by genomic alterations such as CNVs), rather than overall between-group differences, could be involved in the paternal age effect. From the most differentially expressed genes between family 1 and the offspring of young fathers, for example, several show previous literature for their role in aging and psychiatric conditions. *PAFAH1B3* encodes an acetylhydrolase that catalyzes the removal of an acetyl group from the glycerol backbone of platelet-activating factor and shows differentially expression between controls and cases in BD (Nakatani, Hattori et al. 2006) and Williams-Beuron syndrome (Antonell, Vilardell et al. 2010). *PAFAH1B3* has also previously shown reduction in protein levels in brains from aging mice compared to their younger counterparts (Yang, Liu et al. 2008). *PIK3R2* is part of the PI 3-Kinases which are a family of lipid kinases capable of phosphorylating the 3'OH of the inositol ring of phosphoinositides, and has been shown to display age related changes to gene expression in the hippocampus (Zeier, Madorsky et al. 2011). *GSK3A* encodes a multifunctional Ser/Thr protein kinase that is implicated in the control of several regulatory proteins. *GSK3A* shows differentially expression between cases and controls in blood and brain samples in schizophrenia (Glatt, Everall et al. 2005) and is also located in 17q13 in humans which is a candidate region for schizophrenia (Francks, Tozzi et al. 2010). *SLITRK2* encodes an integral membrane protein, and rare variants in the gene have previously been

associated with schizophrenia and autism (Piton, Gauthier et al. 2011), and *SLITRK2* has also been shown to display differential expression in BD (Smith, Bloss et al. 2009).

In family 2, the RAB11FIP5 gene, which is involved in protein trafficking from apical recycling endosomes to the apical plasma membrane, is of interest; a *de novo* balanced translocation in between chromosomes 2 and 9 seen in a patient with pervasive developmental disorder causes the disruption of this locus (Roohi, Tegay et al. 2008). In family 3, *PDE7B* encodes a cAMP-specific phosphodiesterase, a member of the cyclic nucleotide phosphodiesterase family which catalyzes the hydrolysis of cAMP and cGMP to the corresponding 5'-monophosphates and provides a mechanism to down-regulate cAMP and cGMP signalling. *PDE7B* is located in 6q23 which has been previously associated with schizophrenia (Ingason, Giegling et al. 2010). In family 4, *WIF1* is a WNT inhibitory factor, functions as a tumour suppressor gene and is located in 12q14 which is a locus previously associated with autism (Ma, Cuccaro et al. 2007). In family 5, *EPHA5* belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family and has been implicated in mediating developmental events, particularly in the nervous system. *EPHA5* has been implicated as a novel locus in a linkage study of autism (Weiss, Arking et al. 2009).

6.6.4 Limitations

Although we had access to the offspring of three different age fathers, only two age groups were used in the expression analyses presented here. As the third group were from fathers of 12 months old, the differences in expression may be larger in this group but no conclusions about whether the effects are linear can be drawn, although future studies may utilise the available tissue from this group. Due to tissue availability and research funding limitations, RNA from only one tissue was assessed (i.e. frontal cortex), although it would be interesting to examine other tissues such as the cerebellum, in addition to sperm cells from fathers. This makes it difficult to compare the methylation results for the majority of the DMRs assayed in Chapter 5 as only the three DMRs assayed in frontal cortex can be fully compared. QPCR was not carried out to verify the results but is planned for future work. As 18,097 genes were analysed and the uncorrected significance level was set as 0.05, 905 significant differences would be expected by chance which is more than we observed suggesting we observed a lot less or no significant differences between groups.

6.6.5 Conclusions

We do not observe any differences between groups which reach a genome-wide significance value of 2.8×10^{-6} and in light of the issue of multiple testing; all differences observed at a significant level of 0.05 are likely to be by chance. We observed changes in expression in genes associated with inflammatory response and DNA repair which have both been previously highlighted in expression studies of aging as well as genes previously associated with schizophrenia. In family based analysis, some genes which demonstrating group differences in gene expression have previously been associated with aging and psychiatric conditions associated with advanced paternal age. These findings suggest that there are subtle expression changes associated with paternal age in mouse, and that these could be potential candidates for mediating the behavioural differences observed in the offspring of older fathers.

Chapter 7 - Discussion

7.1 Main Findings and Implications

The overall aim of my thesis was to use a rodent model to investigate behavioural and molecular effects associated with advanced paternal age. Previous epidemiological studies, as discussed in Chapter 1, report that increasing paternal age is associated with an increased risk of developing several neuropsychiatric disorders including autism, schizophrenia and BD. In this final chapter I will review the results from my research and relate it back to previous literature, before drawing some general conclusions about the possible underlying mechanisms of the paternal age effect.

7.1.1 Behaviour

In Chapter 3, the main aim was to establish a mouse model and investigate behavioural changes associated with advancing paternal age. The focus of my behavioural analysis was on exploratory behaviour, anxiety related behaviour, short and long term memory and social behaviour, chosen to represent phenotypes related to the neuropsychiatric disorders associated with advanced paternal age in humans. Several behavioural differences were seen in the offspring of older fathers. The offspring of old fathers showed decreased exploration in the holeboard task than in the behaviour of the offspring of young fathers. Compared to the offspring of young fathers, the offspring of old fathers also showed decreased social interaction in the social interaction task in sniffing behaviour and overall behaviour. The offspring of very old fathers showed a reduction in the time spent in overall social behaviour compared with the offspring of young fathers. When the offspring of old and very old fathers were combined into the offspring of fathers over 10 months there was a significant reduction compared to the offspring of young fathers in time spent in sniffing behaviour, time spent allogrooming and time spent in all social behaviours.

When comparing these results to the limited number of other rodent models, it is difficult to draw strong conclusions due to important methodological differences across them (section 3.6.2). The most similar study in terms of methodology is Foldi's *et al's* study (Foldi, Eyles et al. 2010). Their significant results were focussed on females, but our study (like most behavioural analyses) only used male offspring to avoid the confounding effects of hormonal factors. They observed increased exploration in the offspring of the older father group in the elevated plus maze and holeboard task, which is in contrast to the reduction in exploratory behaviour

observed in our study. They also observed no reduction in social behaviour; however the methodology of the task used in their study (as the mice in their study were unable to interact directly) is different to that utilised in my analyses. Overall, the main result from Chapter 3 related to a reduction in social behaviour; taken together with the results from other studies, it can be concluded that there are behavioural changes associated with advancing paternal age although further work is required to specifically pinpoint the specific modality of these changes.

The changes in behaviour reported in Chapter 3 support previous human epidemiological studies which show that regardless of disorder, offspring of older fathers (>45 years) are more likely than the offspring of younger fathers to suffer from social deficits (Weiser, Reichenberg et al. 2008). As social deficits are symptoms of all three disorders which have an increased risk with associated with paternal age (autism, schizophrenia and bipolar disorder), and we observe significant reductions in behaviour, it is reasonable to speculate further that paternal age is related to this common phenotype as opposed to or in addition to increasing risk for the development of these diseases.

7.1.2 Copy Number Variation

In Chapter 4, I explored the role of CNVs as a potential mechanism for the paternal age effect using CGH combined with high-resolution genomic microarrays. Overall, I observed no difference in the frequency, size or type of CNV between the offspring of young fathers and offspring of old fathers. This was true both of both inherited (section 4.5.2.1) and *de novo* CNVs (section 4.5.2.3). Some individuals (primarily breeders) were significant outliers for the overall burden of CNVs, potentially related to lymphoma of the spleen, and were removed from the analyses (section 4.5.1).

Compared to the only other mouse study of CNVs and paternal age, which observed *de novo* CNVs only in their advanced paternal age group (Flatscher-Bader, Foldi et al. 2011), we observed a notable frequency of CNVs in both groups. Although we don't observe a larger number of CNVs in the advanced paternal age group, we observe an overall higher rate of CNVs than in other studies. This could indicate that the mice we used were at a higher risk of

CNVs, that the tissue we used (spleen) is more prone to CNVs, or that there are methodological differences between studies (section 4.6). The occurrence of polymorphic CNVs in populations of inbred (and supposedly genetically identical) mice has implications for the use of these animals as genetic models.

7.1.3 Epigenetic Differences

In Chapter 5, I explored whether there were epigenetic differences (specifically focussing on DNA methylation) between the offspring of young, old and very old fathers. The results in this chapter comprise of three parts. First, in global methylation assayed using LUMA, there were significant increases in methylation with paternal age in spleen and cerebellum compared to the offspring of young fathers (section 5.5.1). Second, looking at DNA methylation levels across DMRs associated with brain-expressed imprinted genes, several DMRs showed paternal age related changes in methylation in cerebellum and three DMRs (Gnas-Nesp, Kcnq1ot1 and Mcts2) were consistently different across the PCR region. All three of these DMRs also showed paternal age related changes to methylation in frontal cortex (section 5.5.2). Third, looking at DNA methylation at the transposable elements IAP and LINE-1, no changes in methylation levels associated with paternal age were observed in any tissue (section 5.5.3).

No previous studies of DNA methylation have been carried out in animal models of advanced paternal age to date, and so our results cannot be compared to any other studies. Although many studies have been carried out on epigenetic changes associated with aging, there are currently no human studies looking at DNA methylation changes associated with paternal age. Imprinted loci were the main focus of this chapter: susceptibility at these loci in terms of methylation has been demonstrated in studies of conception by ART techniques leading to an increase of ASDs (Cox, Burger et al. 2002; DeBaun, Niemitz et al. 2003; Gicquel, Gaston et al. 2003; Maher, Brueton et al. 2003; Orstavik, Eiklid et al. 2003), they are known to be epigenetically labile in the context of environmental exposure, and there is some evidence that they may escape complete epigenetic reprogramming. Our results indicate that certain imprinted loci do show paternal age related changes in methylation, especially at Kcnq1ot1, Gnas-Nesp and Mcts2 as well as at individual sites in other DMRs. In the human ART studies, methylation at *KCNQ1OT1* was shown to be altered in sufferers of Beckwith-Wiedemann

syndrome, a disorder characterised by ASD-related traits (DeBaun, Niemitz et al. 2003; Gicquel, Gaston et al. 2003; Maher, Brueton et al. 2003). This locus also appears to escape complete epigenetic reprogramming in primordial germ cells (Popp, Dean et al. 2010) and so could be a vehicle for the transmission of epigenetic information across generations. It is worth noting that the methylation change at this locus was not observed in our group of offspring from very old fathers; this could be due to the significantly lower group size and the fact that these animals were not produced at the same time as the other two groups with the potential for batch effects. More investigation into this locus is required to determine whether this effect is observed in the offspring of older males, as well as investigation into sperm samples from the actual breeders. By doing this, we could see if any changes which occur in the sperm are transmitted to the offspring or whether the changes occur *de novo*. Focusing on this locus in human disease cohorts as well as the offspring of fathers of different ages would be interesting to further investigate Kcnq1ot1's role in the paternal age effect.

7.1.4 Gene Expression Differences

Chapter 6 aimed to investigate gene expression changes associated with advanced paternal age in the frontal cortex using genome-wide microarrays. We observed over 800 genes showing nominally significant differences between the offspring of young fathers and offspring of old fathers, although the majority of these differences were relatively subtle and none reached a q-value below 0.1 (section 6.5.1.2). Ingenuity pathway analysis of the genes represented in these differences highlighted an enrichment of pathways involved in inflammatory response and DNA replication, recombination and repair (section 6.5.5). No previous studies of gene expression have been carried out on a rodent model of advanced paternal age and so our results are not able to be compared to any similar studies. Human studies of gene expression changes associated with aging, however, have also shown expression changes in genes from pathways associated with inflammatory response, and DNA repair (Lee, Weindruch et al. 2000; Kokkinaki, Lee et al. 2010). Of note, however, we observed no overlapping pathways or genes with a previous human transcriptomic study of advanced paternal age in individuals with autism (Alter, Kharkar et al. 2011). The changes to genes associated with DNA repair are interesting, as increased genomic alterations (e.g. CNVs) obviously represent a key hypothesis for the mediation of the paternal age effect. It would be interesting to investigate expression of these genes in sperm from the breeders as well as looking for genomic alterations in the offspring to see if there are more small alterations

associated with older fathers. Finally for this chapter, we analysed expression changes encompassed within CNVs from Chapter 3. 74% of genes showed a significant reduction in expression when affected by a CNV. The 26% of genes which do not show an expression change could be due to false positives in CNVs, as discussed in section 4.6, or due to the CNV not surrounding the whole gene, with elements such as the transcription factor binding sites not being included in the CNV. However these expression changes provide some further evidence for the legitimacy of the CNV changes we observed.

7.2 Beyond the Paternal Age Effect: Maternal and Grandparental Age

What was not explored in this thesis was the role of advanced maternal age in the increased risk of developing autism, schizophrenia and bipolar disorder. In (*Table 1.1*, *Table 1.2* and *Table 1.3*), it was observed that many epidemiological studies also observe a maternal age effect either in combination with a paternal age effect or individually. Twelve studies report an association between maternal age and ASD, with many identifying a linear association (Grether, Anderson et al. 2009; Golding, Steer et al. 2010), and most studies concluding that a maternal age of over 35 results in a greater risk of the offspring developing ASD. The average age of mothers has also been found to be higher in ASD cases than in controls (Gillberg 1980). In addition to autism, at least seven studies report an association between advanced maternal age and an increased risk of developing other types of neuropsychiatric disease including schizophrenia and other forms of psychosis. For example, the average age of mothers in affected groups was found to be 1-2 years higher than in controls (Hare and Moran 1979; Kinnell 1983), and the percentage of mothers in higher age categories has also been shown to be larger than in control groups (Gillberg 1982; Lee, Malaspina et al. 2011; Naserbakht, Ahmadkhaniha et al. 2011). Interestingly, a recent study looked at the effect of grandparental age and found that a 10-year increment in maternal or paternal grandmothers led to an increased risk of autism (Golding, Steer et al. 2010). This gives speculation that there may be a transgenerational effect of parental age, and this is something that could be addressed in future multi-generational mouse studies.

Maternal age was not investigated in this study, primarily because the literature in support of advanced paternal age's influence on the development of psychiatric conditions is more consistent. Also, because maternal age and paternal age are correlated, it is difficult for many studies to tease apart whether one is more important than the other; a focus on paternal age in our model enabled us to investigate its sole influence. As males can sire offspring naturally for longer than females, there is more potential for a higher paternal age effect to become manifest. Finally, because of the difference in the amount of cell divisions germline cells undergo in males and females (section 1.4.1), and because male germ cells continue to divide throughout the life of males, paternal age was more likely to influence the biological mechanisms explored in this thesis. Building upon this work into the paternal age effect, it would be interesting to explore the maternal age effect using a similar experimental design and also assess parental age in general to see if effects are more pronounced with advanced age in both parents.

7.3 Limitations

There are several limitations to this overall strategy used in this study. First, we only investigated three ages of male breeders and so cannot make any conclusions as to whether the paternal age affect is linear or reaches a plateau. Also the ages do not reach a very old upper limit and the oldest breeders we used were 12 months. Although this corresponds to a decrease in male fertility, with older mice showing a decrease in successful matings (Kidd, Eskenazi et al. 2001), and is past the "retirement" age for commercial male breeders, male mice can be reproductively active for some time after this arbitrary age. The group size of the offspring of very old fathers was substantially smaller than the offspring of young and old fathers. This makes comparing between groups more difficult. Also the breeders from this group were acquired at a different time and so may show a batch effect which could influence the results from this group. Only male offspring were used based upon the sex ratio in autism and the problems of performing behavioural analyses in female mice. However, as reported in section 1.2.2, the sex ratio is actually more even with advancing paternal ages (1.21:1 with a paternal age of over 45) (Anello, Reichenberg et al. 2009), and so we might expect effects to be observed in both males and females. In Foldi *et al's* study of paternal age related behaviour, the main effect they observe is in female offspring of older fathers making it more difficult to compared results across studies. C57BL/6J mice were used in this study as they have previously been shown to perform well in the behavioural tasks selected; we selected

only one strain of mouse in this study to limit the effects of genetic variation on the results. In this regard, however, our CNV data showing considerable between-individual variation is interesting and relevant to the interpretation of future behavioural studies on inbred mice. To look for robustness in the paternal age effects we observed it would be useful to have investigated a second strain for comparison. Although we investigated DNA methylation levels in sperm from different age males and looked for *de novo* CNVs using spleen tissue from the breeders used, it would have been more informative to have used sperm samples from the breeders themselves. First it would have enabled us to look for epigenetic inheritance by using DNA obtained from sperm of the actual breeders used, and second, we could have drawn more conclusions about whether the *de novo* CNVs observed in the offspring originated in the sperm, or occurred post-conception. The limitations of the available technologies are also something which should be addressed. As the cost of running genome-wide molecular approaches decreases, the methods within our means to use will become greater. For example, future studies will utilise RNA-seq and a more genome-wide approach for DNA methylation analysis (such as whole genome bisulfite sequencing) for more in depth analysis. We used genome wide approached in Chapter 4 (CNVs) and Chapter 6 (expression), however due to time constraints, verification of array data is not yet completed, however analysis is ongoing.

7.4 Overall Conclusions

From the research carried out in this thesis, it can be concluded that there is some evidence for behavioural changes in the offspring associated with advanced paternal age, and interestingly they are mainly seen in social behaviour, reflecting deficits common to autism, schizophrenia and bipolar disorder. Genome-wide analysis of structural genomic variation suggests that the offspring of older fathers are not characterised by an increased burden of CNVs, although I did observe some evidence for altered DNA methylation and subtle changes in gene expression. Whether these changes occur in the sperm of the fathers or *de novo* in the offspring however, is not clear. It is also plausible that DNA methylation changes result from point mutation in CpG rich areas, which were not investigated in this study.

7.5 Future Work

It would be interesting to continue this work in many ways. First by assessing the DNA methylation and transcriptomic differences we observed in our mouse samples in human samples to investigate whether they are related to paternal age or psychiatric condition in humans. Second, it is important to pursue the animal research further, especially by using female as well as male offspring to see if the behavioural and molecular changes occur in both sexes. By increasing the range of ages of the fathers and taking a more quantitative approach we could see i) if there is a continuous effect of paternal age, ii) if the effect reaches a plateau or iii) whether there is a specific 'at risk' parental age group. In some human studies, the offspring of very young fathers also show an increased risk of developing psychiatric disorders, and this would be interesting to also assess in a rodent model. It would also be interesting to use breeders for multiple litters at different ages to see if offspring produced from earlier matings showed differences from the offspring of matings carried out later with the same father. Furthermore, as discussed above, it would be informative to also examine the effects of advanced maternal age, and extend our analyses to other strains of mouse. Finally, the use of methods to isolate sperm from males of different ages would enable the analysis of chromosomal abnormalities and methylation differences occurring in the father, to establish the rate of molecular changes associated with advanced paternal age.

7.6 Final Word

To date, most research into the effect of paternal age in the development of psychiatric disease has been epidemiological in nature and this study represents one of the first attempts to model this in rodents, examining both behavioural and molecular changes induced by advanced paternal age. As the age at which people are having children increases, the impact of paternal age on the risk for psychiatric disorders is likely to become more pronounced, highlighting the need for more research into this area. With more animal and human studies working towards finding a mechanism for the effect, it is my hope that we will be able to use this knowledge not only in understanding the effect of paternal age on the offspring, but also the aetiology of the disorders involved.

Bibliography

(2008). Partek® Genomics SuiteTM. St. Louis, MO, USA, Partek Inc.

Abdolmaleky, H. M., K. H. Cheng, et al. (2006). "Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder." Hum Mol Genet **15**(21): 3132-45.

Abdolmaleky, H. M., K. H. Cheng, et al. (2005). "Hypermethylation of the reelin (RELN) promoter in the brain of schizophrenic patients: a preliminary report." Am J Med Genet B Neuropsychiatr Genet **134B**(1): 60-6.

Adkins, R. M., G. Somes, et al. (2010). "Association of birth weight with polymorphisms in the IGF2, H19, and IGF2R genes." Pediatr Res **68**(5): 429-34.

Adkins, R. M., F. Thomas, et al. (2011). "Parental ages and levels of DNA methylation in the newborn are correlated." BMC Med Genet **12**: 47.

Agam, A., B. Yalcin, et al. (2010). "Elusive copy number variation in the mouse genome." PLoS One **5**(9): e12839.

Alaerts, M. and J. Del-Favero (2009). "Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future." Hum Mutat **30**(8): 1139-52.

Aleman, A., R. S. Kahn, et al. (2003). "Sex differences in the risk of schizophrenia: evidence from meta-analysis." Arch Gen Psychiatry **60**(6): 565-71.

Alloy, L. B., L. Y. Abramson, et al. (2005). "The psychosocial context of bipolar disorder: environmental, cognitive, and developmental risk factors." Clin Psychol Rev **25**(8): 1043-75.

Altar, C. A., L. W. Jurata, et al. (2005). "Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts." Biol Psychiatry **58**(2): 85-96.

Alter, M. D., R. Kharkar, et al. (2011). "Autism and increased paternal age related changes in global levels of gene expression regulation." PLoS One **6**(2): e16715.

American-Psychiatric-Association (2000). Diagnostic and statistical manual of mental disorders. Washington D.C.

Andreasson, S., P. Allebeck, et al. (1987). "Cannabis and schizophrenia. A longitudinal study of Swedish conscripts." Lancet **2**(8574): 1483-6.

Anello, A., A. Reichenberg, et al. (2009). "Brief report: parental age and the sex ratio in autism." J Autism Dev Disord **39**(10): 1487-92.

Antonell, A., M. Vilardell, et al. (2010). "Transcriptome profile in Williams-Beuren syndrome lymphoblast cells reveals gene pathways implicated in glucose intolerance and visuospatial construction deficits." Hum Genet **128**(1): 27-37.

Anway, M. D., A. S. Cupp, et al. (2005). "Epigenetic transgenerational actions of endocrine disruptors and male fertility." Science **308**(5727): 1466-9.

Ariel, M., H. Cedar, et al. (1994). "Developmental changes in methylation of spermatogenesis-specific genes include reprogramming in the epididymis." Nat Genet **7**(1): 59-63.

Arita, A. and M. Costa (2011). Environmental Agents and Epigenetics. Handbook of Epigenetics. T. Tollefsbol, Elsevier Inc. **1:** 459-476.

Asherson, P., C. Walsh, et al. (1994). "Imprinting and anticipation. Are they relevant to genetic studies of schizophrenia?" Br J Psychiatry **164**(5): 619-24.

Auroux, M. (1983). "Decrease of learning capacity in offspring with increasing paternal age in the rat." Teratology **27**(2): 141-8.

Auroux, M. R., M. J. Mayaux, et al. (1989). "Paternal age and mental functions of progeny in man." Hum Reprod **4**(7): 794-7.

Avner, P. and E. Heard (2001). "X-chromosome inactivation: counting, choice and initiation." Nat Rev Genet **2**(1): 59-67.

Badcock, C. and B. Crespi (2006). "Imbalanced genomic imprinting in brain development: an evolutionary basis for the aetiology of autism." J Evol Biol **19**(4): 1007-32.

Baillie, J. K., M. W. Barnett, et al. (2011). "Somatic retrotransposition alters the genetic landscape of the human brain." Nature **479**(7374): 534-7.

Baron, C. A., S. Y. Liu, et al. (2006). "Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism." J Autism Dev Disord **36**(8): 973-82.

Beaudet, A. L. (2007). "Autism: highly heritable but not inherited." Nat Med **13**(5): 534-6.

Bell, J. T., A. A. Pai, et al. (2011). "DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines." Genome Biol **12**(1): R10.

Berg, D., C. Holzmann, et al. (2003). "14-3-3 proteins in the nervous system." Nat Rev Neurosci **4**(9): 752-62.

Berg, J. S., N. Brunetti-Pierri, et al. (2007). "Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region." Genet Med **9**(7): 427-41.

Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**(7143): 407-12.

Bertram, L., R. Busch, et al. (1998). "Paternal age is a risk factor for Alzheimer disease in the absence of a major gene." Neurogenetics **1**(4): 277-80.

Bezchlibnyk, Y. B., J. F. Wang, et al. (2001). "Gene expression differences in bipolar disorder revealed by cDNA array analysis of post-mortem frontal cortex." J Neurochem **79**(4): 826-34.

Biermann, K. and K. Steger (2007). "Epigenetics in male germ cells." J Androl **28**(4): 466-80.

Biliya, S. and L. A. Bulla, Jr. (2010). "Genomic imprinting: the influence of differential methylation in the two sexes." Exp Biol Med (Maywood) **235**(2): 139-47.

Bird, A. P. (1980). "DNA methylation and the frequency of CpG in animal DNA." Nucleic Acids Res **8**(7): 1499-504.

Bird, A. P. and A. P. Wolffe (1999). "Methylation-induced repression--belts, braces, and chromatin." Cell **99**(5): 451-4.

Bjornsson, H. T., M. I. Sigurdsson, et al. (2008). "Intra-individual change over time in DNA methylation with familial clustering." JAMA **299**(24): 2877-83.

Blackwood, D. H., A. Fordyce, et al. (2001). "Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family." Am J Hum Genet **69**(2): 428-33.

Blankenberg, D., G. Von Kuster, et al. (2010). "Galaxy: a web-based genome analysis tool for experimentalists." Curr Protoc Mol Biol **Chapter 19**: Unit 19 10 1-21.

Bocklandt, S., W. Lin, et al. (2011). "Epigenetic predictor of age." PLoS One **6**(6): e14821.

Boissier, J. R. and P. Simon (1962). "[The exploration reaction in the mouse. Preliminary note]." Therapie **17**: 1225-32.

Boissier, J. R., P. Simon, et al. (1964). "[Use of a Particular Mouse Reaction (Hole Board Method) for the Study of Psychotropic Drugs]." Therapie **19**: 571-83.

Boks, M. P., E. M. Derks, et al. (2009). "The relationship of DNA methylation with age, gender and genotype in twins and healthy controls." PLoS One **4**(8): e6767.

Bork, S., S. Pfister, et al. (2010). "DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells." Aging Cell **9**(1): 54-63.

Brambilla, P., A. Hardan, et al. (2003). "Brain anatomy and development in autism: review of structural MRI studies." Brain Res Bull **61**(6): 557-69.

Brayton, C. F., P. M. Treuting, et al. (2012). "Pathobiology of Aging Mice and GEM: Background Strains and Experimental Design." Vet Pathol **49**(1): 85-105.

Breart, G. (1997). "Delayed childbearing." Eur J Obstet Gynecol Reprod Biol **75**(1): 71-3.

Brown, A. S., C. A. Schaefer, et al. (2002). "Paternal age and risk of schizophrenia in adult offspring." Am J Psychiatry **159**(9): 1528-33.

Bruce, H. A., N. Sachs, et al. (2009). "Long tandem repeats as a form of genomic copy number variation: structure and length polymorphism of a chromosome 5p repeat in control and schizophrenia populations." Psychiatr Genet **19**(2): 64-71.

Bruder, C. E., A. Piotrowski, et al. (2008). "Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles." Am J Hum Genet **82**(3): 763-71.

Buizer-Voskamp, J. E., W. Laan, et al. (2011). "Paternal age and psychiatric disorders: Findings from a Dutch population registry." Schizophr Res.

Byrne, M., E. Agerbo, et al. (2003). "Parental age and risk of schizophrenia: a case-control study." Arch Gen Psychiatry **60**(7): 673-8.

Cahan, P., Y. Li, et al. (2009). "The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells." Nat Genet **41**(4): 430-7.

Cantor, R. M., J. L. Yoon, et al. (2007). "Paternal age and autism are associated in a family-based sample." Mol Psychiatry **12**(5): 419-21.

Cardona Maya, W., J. Berdugo, et al. (2009). "The effects of male age on semen parameters: analysis of 1364 men attending an andrology center." Aging Male **12**(4): 100-3.

Cardwell, C. R., L. C. Stene, et al. (2010). "Maternal age at birth and childhood type 1 diabetes: a pooled analysis of 30 observational studies." Diabetes **59**(2): 486-94.

Carper, R. A., P. Moses, et al. (2002). "Cerebral lobes in autism: early hyperplasia and abnormal age effects." Neuroimage **16**(4): 1038-51.

Carrard, A., A. Salzmann, et al. (2011). "Increased DNA methylation status of the serotonin receptor 5HTR1A gene promoter in schizophrenia and bipolar disorder." J Affect Disord **132**(3): 450-3.

Cassidy, S. B. (1997). "Prader-Willi syndrome." J Med Genet **34**(11): 917-23.

Centola, G. M. and S. Eberly (1999). "Seasonal variations and age-related changes in human sperm count, motility, motion parameters, morphology, and white blood cell concentration." Fertil Steril **72**(5): 803-8.

Chandley, A. C., P. Edmond, et al. (1975). "Cytogenetics and infertility in man. I. Karyotype and seminal analysis: results of a five-year survey of men attending a subfertility clinic." Ann Hum Genet **39**(2): 231-54.

Cheng, Z., M. Ventura, et al. (2005). "A genome-wide comparison of recent chimpanzee and human segmental duplications." Nature **437**(7055): 88-93.

Choudary, P. V., M. Molnar, et al. (2005). "Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression." Proc Natl Acad Sci U S A **102**(43): 15653-8.

Christensen, B. C., E. A. Houseman, et al. (2009). "Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context." PLoS Genet **5**(8): e1000602.

Chubb, J. E., N. J. Bradshaw, et al. (2008). "The DISC locus in psychiatric illness." Mol Psychiatry **13**(1): 36-64.

Clark, S. J., J. Harrison, et al. (1994). "High sensitivity mapping of methylated cytosines." Nucleic Acids Res **22**(15): 2990-7.

Clayton-Smith, J. and M. E. Pembrey (1992). "Angelman syndrome." J Med Genet **29**(6): 412-5.

Colangelo, V., J. Schurr, et al. (2002). "Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling." J Neurosci Res **70**(3): 462-73.

Conrad, D. F., T. D. Andrews, et al. (2006). "A high-resolution survey of deletion polymorphism in the human genome." Nat Genet **38**(1): 75-81.

Cook, E. H., Jr., R. Y. Courchesne, et al. (1998). "Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers." Am J Hum Genet **62**(5): 1077-83.

Cook, E. H., Jr., V. Lindgren, et al. (1997). "Autism or atypical autism in maternally but not paternally derived proximal 15q duplication." Am J Hum Genet **60**(4): 928-34.

Coulondre, C., J. H. Miller, et al. (1978). "Molecular basis of base substitution hotspots in Escherichia coli." Nature **274**(5673): 775-80.

Court, J., D. Spurden, et al. (1999). "Neuronal nicotinic receptors in dementia with Lewy bodies and schizophrenia: alpha-bungarotoxin and nicotine binding in the thalamus." J Neurochem **73**(4): 1590-7.

Cox, G. F., J. Burger, et al. (2002). "Intracytoplasmic sperm injection may increase the risk of imprinting defects." Am J Hum Genet **71**(1): 162-4.

Croen, L. A., J. K. Grether, et al. (2002). "Descriptive epidemiology of autism in a California population: who is at risk?" J Autism Dev Disord **32**(3): 217-24.

Croen, L. A., D. V. Najjar, et al. (2007). "Maternal and paternal age and risk of autism spectrum disorders." Arch Pediatr Adolesc Med **161**(4): 334-40.

Crouse, H. V. (1960). "The Controlling Element in Sex Chromosome Behavior in Sciara." Genetics **45**(10): 1429-43.

Crow, J. F. (1999). "Spontaneous mutation in man." Mutat Res **437**(1): 5-9.

Crow, J. F. (2000). "The origins, patterns and implications of human spontaneous mutation." Nat Rev Genet **1**(1): 40-7.

Crow, T. J., L. E. DeLisi, et al. (1989). "Concordance by sex in sibling pairs with schizophrenia is paternally inherited. Evidence for a pseudoautosomal locus." Br J Psychiatry **155**: 92-7.

Cui, H., M. Cruz-Correa, et al. (2003). "Loss of IGF2 imprinting: a potential marker of colorectal cancer risk." Science **299**(5613): 1753-5.

Curley, J. P., F. A. Champagne, et al. (2008). "Transgenerational effects of impaired maternal care on behaviour of offspring and grandoffspring." Animal Behaviour **75**(4): 1551-1561.

Curley, J. P., R. Mashoodh, et al. (2010). "Epigenetics and the origins of paternal effects." Horm Behav **59**(3): 306-14.

Cutler, G., L. A. Marshall, et al. (2007). "Significant gene content variation characterizes the genomes of inbred mouse strains." Genome Res **17**(12): 1743-54.

Dalen, P., M. L. Dahl, et al. (1998). "10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes." Clin Pharmacol Ther **63**(4): 444-52.

Davies, M. N., M. Volta, et al. (Under Review). "Tissue-specific epigenetic variation across brain and blood: functional annotation of the human brain methylome." Neuron.

Davies, W., A. R. Isles, et al. (2005). "Imprinted gene expression in the brain." Neurosci Biobehav Rev **29**(3): 421-30.

Daxinger, L. and E. Whitelaw (2010). "Transgenerational epigenetic inheritance: more questions than answers." Genome Res **20**(12): 1623-8.

de Boer, P., L. Ramos, et al. (2010). "Memoirs of an insult: sperm as a possible source of transgenerational epimutations and genetic instability." Mol Hum Reprod **16**(1): 48-56.

de la Rochebrochard, E. and P. Thonneau (2002). "Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study." Hum Reprod **17**(6): 1649-56.

DeBaun, M. R., E. L. Niemitz, et al. (2003). "Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19." Am J Hum Genet **72**(1): 156-60.

Deelen, J., H. W. Uh, et al. (2011). "Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways." Age (Dordr).

Deininger, P. L., J. V. Moran, et al. (2003). "Mobile elements and mammalian genome evolution." Curr Opin Genet Dev **13**(6): 651-8.

Dempster, E. L., R. Pidsley, et al. (2011). "Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder." Hum Mol Genet.

Diskin, S. J., M. Li, et al. (2008). "Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms." Nucleic Acids Res **36**(19): e126.

Dolinoy, D. C., D. Huang, et al. (2007). "Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development." Proc Natl Acad Sci U S A **104**(32): 13056-61.

Dolinoy, D. C. and R. L. Jirtle (2008). "Environmental epigenomics in human health and disease." Environ Mol Mutagen **49**(1): 4-8.

Doornbos, M., B. Sikkema-Raddatz, et al. (2009). "Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances." Eur J Med Genet **52**(2-3): 108-15.

Drake, J. W., B. Charlesworth, et al. (1998). "Rates of spontaneous mutation." Genetics **148**(4): 1667-86.

Du, P., W. A. Kibbe, et al. (2007). "nuID: a universal naming scheme of oligonucleotides for illumina, affymetrix, and other microarrays." Biol Direct **2**: 16.

Du, P., W. A. Kibbe, et al. (2008). "lumi: a pipeline for processing Illumina microarray." Bioinformatics **24**(13): 1547-8.

Durkin, M. S., M. J. Maenner, et al. (2008). "Advanced parental age and the risk of autism spectrum disorder." Am J Epidemiol **168**(11): 1268-76.

Eastwood, S. L., R. W. Kerwin, et al. (1997). "Immunoautoradiographic evidence for a loss of alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate-preferring non-N-methyl-D-aspartate glutamate receptors within the medial temporal lobe in schizophrenia." Biol Psychiatry **41**(6): 636-43.

Eastwood, S. L., B. McDonald, et al. (1995). "Decreased expression of mRNAs encoding non-NMDA glutamate receptors GluR1 and GluR2 in medial temporal lobe neurons in schizophrenia." Brain Res Mol Brain Res **29**(2): 211-23.

Egan, C. M., S. Sridhar, et al. (2007). "Recurrent DNA copy number variation in the laboratory mouse." Nat Genet **39**(11): 1384-9.

Eickbush, T. H. and A. V. Furano (2002). "Fruit flies and humans respond differently to retrotransposons." Curr Opin Genet Dev **12**(6): 669-74.

El-Saadi, O., C. B. Pedersen, et al. (2004). "Paternal and maternal age as risk factors for psychosis: findings from Denmark, Sweden and Australia." Schizophr Res **67**(2-3): 227-36.

Elliott, B. and M. Jasin (2002). "Double-strand breaks and translocations in cancer." Cell Mol Life Sci **59**(2): 373-85.

Emamian, E. S., D. Hall, et al. (2004). "Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia." Nat Genet **36**(2): 131-7.

Engemann, S., M. Strodicke, et al. (2000). "Sequence and functional comparison in the Beckwith-Wiedemann region: implications for a novel imprinting centre and extended imprinting." Hum Mol Genet **9**(18): 2691-706.

Ennaceur, A. and J. Delacour (1988). "A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data." Behav Brain Res **31**(1): 47-59.

Eskenazi, B., A. J. Wyrobek, et al. (2003). "The association of age and semen quality in healthy men." Hum Reprod **18**(2): 447-54.

EuropeanBioinformaticsInstitute. (2011). "European Bioinformatics Institute." from http://www.ebi.ac.uk/gxa/gene/ENSMUSG00000042814?ef=organism_part.

Evenson, D. P., L. K. Jost, et al. (1999). "Utility of the sperm chromatin structure assay as a diagnostic and prognostic tool in the human fertility clinic." Hum Reprod **14**(4): 1039-49.

Fang, M. Z., Y. Wang, et al. (2003). "Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines." Cancer Res **63**(22): 7563-70.

Fatemi, S. H., A. V. Snow, et al. (2005). "Reelin signaling is impaired in autism." Biol Psychiatry **57**(7): 777-87.

Fernandez, B. A., W. Roberts, et al. (2010). "Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder." J Med Genet **47**(3): 195-203.

Fernandez, T., T. Morgan, et al. (2008). "Disruption of Contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome." Am J Hum Genet **82**(6): 1385.

Flanagan, J. M., V. Popendikyte, et al. (2006). "Intra- and interindividual epigenetic variation in human germ cells." Am J Hum Genet **79**(1): 67-84.

Flatscher-Bader, T., C. Foldi, et al. (2011). "Increased de novo copy number variants in the offspring of older males." Translational Psychiatry.

Foldi, C. J., D. W. Eyles, et al. (2010). "Advanced paternal age is associated with alterations in discrete behavioural domains and cortical neuroanatomy of C57BL/6J mice." Eur J Neurosci **31**(3): 556-64.

Fombonne, E. (2003). "Epidemiological surveys of autism and other pervasive developmental disorders: an update." J Autism Dev Disord **33**(4): 365-82.

Fowden, A. L., C. Sibley, et al. (2006). "Imprinted genes, placental development and fetal growth." Horm Res **65 Suppl 3**: 50-8.

Fraga, M. F., E. Ballestar, et al. (2005). "Epigenetic differences arise during the lifetime of monozygotic twins." Proc Natl Acad Sci U S A **102**(30): 10604-9.

Fraga, M. F. and M. Esteller (2007). "Epigenetics and aging: the targets and the marks." Trends Genet **23**(8): 413-8.

Francks, C., F. Tozzi, et al. (2010). "Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13." Mol Psychiatry **15**(3): 319-25.

Frans, E. M., S. Sandin, et al. (2008). "Advancing paternal age and bipolar disorder." Arch Gen Psychiatry **65**(9): 1034-40.

Freedman, R., H. Coon, et al. (1997). "Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus." Proc Natl Acad Sci U S A **94**(2): 587-92.

Freedman, R., M. Hall, et al. (1995). "Evidence in postmortem brain tissue for decreased numbers of hippocampal nicotinic receptors in schizophrenia." Biol Psychiatry **38**(1): 22-33.

Freeman, J. L., G. H. Perry, et al. (2006). "Copy number variation: new insights in genome diversity." Genome Res **16**(8): 949-61.

Friedman, J. I., T. Vrijenhoek, et al. (2008). "CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy." Mol Psychiatry **13**(3): 261-6.

Gabis, L., R. Raz, et al. (2010). "Paternal age in autism spectrum disorders and ADHD." Pediatr Neurol **43**(4): 300-2.

Garcia-Palomares, S., J. F. Pertusa, et al. (2009). "Long-term effects of delayed fatherhood in mice on postnatal development and behavioral traits of offspring." Biol Reprod **80**(2): 337-42.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-82.

Ghaziuddin, M. and M. Burmeister (1999). "Deletion of chromosome 2q37 and autism: a distinct subtype?" J Autism Dev Disord **29**(3): 259-63.

Gicquel, C., V. Gaston, et al. (2003). "In vitro fertilization may increase the risk of Beckwith-Wiedemann syndrome related to the abnormal imprinting of the KCN1OT gene." Am J Hum Genet **72**(5): 1338-41.

Gillberg, C. (1980). "Maternal age and infantile autism." J Autism Dev Disord **10**(3): 293-7.

Gillberg, C. (1982). "Parental age in child psychiatric clinic attenders." Acta Psychiatr Scand **66**(6): 471-8.

Glasson, E. J., C. Bower, et al. (2004). "Perinatal factors and the development of autism: a population study." Arch Gen Psychiatry **61**(6): 618-27.

Glatt, S. J., I. P. Everall, et al. (2005). "Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia." Proc Natl Acad Sci U S A **102**(43): 15533-8.

Glessner, J. T., M. P. Reilly, et al. (2010). "Strong synaptic transmission impact by copy number variations in schizophrenia." Proc Natl Acad Sci U S A **107**(23): 10584-9.

Glessner, J. T., K. Wang, et al. (2009). "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes." Nature **459**(7246): 569-73.

Goecks, J., A. Nekrutenko, et al. (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." Genome Biol **11**(8): R86.

Golding, J., C. Steer, et al. (2010). "Parental and grandparental ages in the autistic spectrum disorders: a birth cohort study." PLoS One **5**(4): e9939.

Goll, M. G., F. Kirpekar, et al. (2006). "Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2." Science **311**(5759): 395-8.

Gonzalez, E., H. Kulkarni, et al. (2005). "The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility." Science **307**(5714): 1434-40.

Graubert, T. A., P. Cahan, et al. (2007). "A high-resolution map of segmental DNA copy number variation in the mouse genome." PLoS Genet **3**(1): e3.

Grayson, D. R., X. Jia, et al. (2005). "Reelin promoter hypermethylation in schizophrenia." Proc Natl Acad Sci U S A **102**(26): 9341-6.

Gregg, C., J. Zhang, et al. (2010). "High-resolution analysis of parent-of-origin allelic expression in the mouse brain." Science **329**(5992): 643-8.

Gregg, J. P., L. Lit, et al. (2008). "Gene expression changes in children with autism." Genomics **91**(1): 22-9.

Gregory, S. G., J. J. Connelly, et al. (2009). "Genomic and epigenetic evidence for oxytocin receptor deficiency in autism." BMC Med **7**: 62.

Grether, J. K., M. C. Anderson, et al. (2009). "Risk of autism and increasing maternal and paternal age in a large north American population." Am J Epidemiol **170**(9): 1118-26.

Grozeva, D., G. Kirov, et al. (2010). "Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia." Arch Gen Psychiatry **67**(4): 318-27.

Guan, Z. Z., X. Zhang, et al. (1999). "Decreased protein level of nicotinic receptor alpha7 subunit in the frontal cortex from schizophrenic brain." Neuroreport **10**(8): 1779-82.

Haig, D. and C. Graham (1991). "Genomic imprinting and the strange case of the insulin-like growth factor II receptor." Cell **64**(6): 1045-6.

Hall, C. S. (1951). The Genetics of Behaviour. Handbook of Experimental Psychology. S. S. Steven. New York, John Wiley & Sons**:** 304-329.

Hammiche, F., J. S. Laven, et al. (2010). "Sperm quality decline among men below 60 years of age undergoing IVF or ICSI treatment." J Androl **32**(1): 70-6.

Han, H., C. C. Cortez, et al. (2011). "DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter." Hum Mol Genet **20**(22): 4299-310.

Hare, E. H. and P. A. Moran (1979). "Raised parental age in psychiatric patients: evidence for the constitutional hypothesis." Br J Psychiatry **134**: 169-77.

Harrison, P. J. (1999). "Neurochemical alterations in schizophrenia affecting the putative receptor targets of atypical antipsychotics. Focus on dopamine (D1, D3, D4) and 5-HT2a receptors." Br J Psychiatry Suppl(38): 12-22.

Harrison, P. J., N. Freemantle, et al. (2003). "Meta-analysis of brain weight in schizophrenia." Schizophr Res **64**(1): 25-34.

Hashimoto, R., R. E. Straub, et al. (2004). "Expression analysis of neuregulin-1 in the dorsolateral prefrontal cortex in schizophrenia." Mol Psychiatry **9**(3): 299-307.

Hastings, P. J., J. R. Lupski, et al. (2009). "Mechanisms of change in gene copy number." Nat Rev Genet **10**(8): 551-64.

Heck, K. E., K. C. Schoendorf, et al. (1997). "Delayed childbearing by education level in the United States, 1969-1994." Matern Child Health J **1**(2): 81-8.

Hehir-Kwa, J. Y., B. Rodriguez-Santiago, et al. (2011). "De novo copy number variants associated with intellectual disability have a paternal origin and age bias." J Med Genet **48**(11): 776-8.

Heidmann, O. and T. Heidmann (1991). "Retrotransposition of a mouse IAP sequence tagged with an indicator gene." Cell **64**(1): 159-70.

Heijmans, B. T., D. Kremer, et al. (2007). "Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus." Hum Mol Genet **16**(5): 547-54.

Heijmans, B. T., E. W. Tobi, et al. (2008). "Persistent epigenetic differences associated with prenatal exposure to famine in humans." Proc Natl Acad Sci U S A **105**(44): 17046-9.

Hendrich, B. and S. Tweedie (2003). "The methyl-CpG binding domain and the evolving role of DNA methylation in animals." Trends Genet **19**(5): 269-77.

Henikoff, S. and M. A. Matzke (1997). "Exploring and explaining epigenetic effects." Trends Genet **13**(8): 293-5.

Henrichsen, C. N., N. Vinckenbosch, et al. (2009). "Segmental copy number variation shapes tissue transcriptomes." Nat Genet **41**(4): 424-9.

Hernandez, D. G., M. A. Nalls, et al. (2011). "Distinct DNA methylation changes highly correlated with chronological age in the human brain." Hum Mol Genet **20**(6): 1164-72.

Hook, E. B. (1976). "Estimates of maternal age-specific risks of Down-syndrome birth in women aged 34-41." Lancet **2**(7975): 33-4.

Hu, V. W., B. C. Frank, et al. (2006). "Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes." BMC Genomics **7**: 118.

Huang, L., R. Sauve, et al. (2008). "Maternal age and risk of stillbirth: a systematic review." CMAJ **178**(2): 165-72.

Huether, C. A., J. Ivanovich, et al. (1998). "Maternal age specific risk rate estimates for Down syndrome among live births in whites and other races from Ohio and metropolitan Atlanta, 1970-1989." J Med Genet **35**(6): 482-90.

Hultman, C. M., S. Sandin, et al. (2010). "Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies." Mol Psychiatry.

Hultman, C. M., P. Sparen, et al. (2002). "Perinatal risk factors for infantile autism." Epidemiology **13**(4): 417-23.

Iafrate, A. J., L. Feuk, et al. (2004). "Detection of large-scale variation in the human genome." Nat Genet **36**(9): 949-51.

Ikeda, M., B. Aleksic, et al. (2010). "Copy number variation in schizophrenia in the Japanese population." Biol Psychiatry **67**(3): 283-6.

Ingason, A., I. Giegling, et al. (2010). "A large replication study and meta-analysis in European samples provides further support for association of AHI1 markers with schizophrenia." Hum Mol Genet **19**(7): 1379-86.

Ingason, A., G. Kirov, et al. (2011). "Maternally derived microduplications at 15q11-q13: implication of imprinted genes in psychotic illness." Am J Psychiatry **168**(4): 408-17.

Ingason, A., D. Rujescu, et al. (2011). "Copy number variations of chromosome 16p13.1 region associated with schizophrenia." Mol Psychiatry **16**(1): 17-25.

Inoue, K. and J. R. Lupski (2002). "Molecular mechanisms for genomic disorders." Annu Rev Genomics Hum Genet **3**: 199-242.

International Schizophrenia Consortium, T. (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." Nature **455**(7210): 237-41.

Irizarry, R. A., C. Ladd-Acosta, et al. (2009). "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." Nat Genet **41**(2): 178-86.

Issa, J. P., P. M. Vertino, et al. (1996). "Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis." Proc Natl Acad Sci U S A **93**(21): 11757-62.

Itsara, A., G. M. Cooper, et al. (2009). "Population analysis of large copy number variants and hotspots of human genetic disease." Am J Hum Genet **84**(2): 148-61.

Itsara, A., H. Wu, et al. (2010). "De novo rates and selection of large copy number variation." Genome Res **20**(11): 1469-81.

Iwamoto, K., M. Bundo, et al. (2005). "DNA methylation status of SOX10 correlates with its downregulation and oligodendrocyte dysfunction in schizophrenia." J Neurosci **25**(22): 5376-81.

Iwamoto, K., C. Kakiuchi, et al. (2004). "Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders." Mol Psychiatry **9**(4): 406-16.

Jacobsson, B., L. Ladfors, et al. (2004). "Advanced maternal age and adverse perinatal outcome." Obstet Gynecol **104**(4): 727-33.

Jacquemont, M. L., D. Sanlaville, et al. (2006). "Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders." J Med Genet **43**(11): 843-9.

Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nat Genet **33 Suppl**: 245-54.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.

Jirtle, R. L. (2006). "geneimprint." from http://www.geneimprint.com/.

Jirtle, R. L., M. Sander, et al. (2000). "Genomic imprinting and environmental disease susceptibility." Environ Health Perspect **108**(3): 271-8.

Jirtle, R. L. and M. K. Skinner (2007). "Environmental epigenomics and disease susceptibility." Nat Rev Genet **8**(4): 253-62.

Johnson, K. J., S. E. Carozza, et al. (2009). "Parental age and risk of childhood cancer: a pooled analysis." Epidemiology **20**(4): 475-83.

Johnson, W. E., C. Li, et al. (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods." Biostatistics **8**(1): 118-27.

Johnston-Wilson, N. L., C. D. Sims, et al. (2000). "Disease-specific alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder. The Stanley Neuropathology Consortium." Mol Psychiatry **5**(2): 142-9.

Jung, A., H. C. Schuppe, et al. (2002). "Comparison of semen quality in older and younger men attending an andrology clinic." Andrologia **34**(2): 116-22.

Jurata, L. W., Y. V. Bukhman, et al. (2004). "Comparison of microarray-based mRNA profiling technologies for identification of psychiatric disease and drug signatures." J Neurosci Methods **138**(1-2): 173-88.

Kafri, T., M. Ariel, et al. (1992). "Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line." Genes Dev **6**(5): 705-14.

Kaminsky, Z. A., T. Tang, et al. (2009). "DNA methylation profiles in monozygotic and dizygotic twins." Nat Genet **41**(2): 240-5.

Kamnasaran, D., W. J. Muir, et al. (2003). "Disruption of the neuronal PAS3 gene in a family affected with schizophrenia." J Med Genet **40**(5): 325-32.

Kanner, L. (1968). "Autistic disturbances of affective contact." Acta Paedopsychiatr **35**(4): 100-36.

Karimi, M., S. Johansson, et al. (2006). "Using LUMA: a Luminometric-based assay for global DNA-methylation." Epigenetics **1**(1): 45-8.

Karimi, M., S. Johansson, et al. (2006). "LUMA (LUminometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation." Exp Cell Res **312**(11): 1989-95.

Kazazian, H. H., Jr. and J. V. Moran (1998). "The impact of L1 retrotransposons on the human genome." Nat Genet **19**(1): 19-24.

Kemper, T. L. and M. Bauman (1998). "Neuropathology of infantile autism." J Neuropathol Exp Neurol **57**(7): 645-52.

Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kerkel, K., A. Spadola, et al. (2008). "Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation." Nat Genet **40**(7): 904-8.

Kermicle, J. L. (1970). "Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission." Genetics **66**(1): 69-85.

Kidd, J. M., G. M. Cooper, et al. (2008). "Mapping and sequencing of structural variation from eight human genomes." Nature **453**(7191): 56-64.

Kidd, S. A., B. Eskenazi, et al. (2001). "Effects of male age on semen quality and fertility: a review of the literature." Fertil Steril **75**(2): 237-48.

Kim, H. G., S. Kishikawa, et al. (2008). "Disruption of neurexin 1 associated with autism spectrum disorder." Am J Hum Genet **82**(1): 199-207.

Kim, J. K., M. Samaranayake, et al. (2009). "Epigenetic mechanisms in mammals." Cell Mol Life Sci **66**(4): 596-612.

Kim, Y. K., A. M. Myint, et al. (2004). "Th1, Th2 and Th3 cytokine alteration in schizophrenia." Prog Neuropsychopharmacol Biol Psychiatry **28**(7): 1129-34.

Kinnell, H. G. (1983). "Parental age in schizophrenia." Br J Psychiatry **142**: 204.

Kirov, G., D. Gumus, et al. (2008). "Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia." Hum Mol Genet **17**(3): 458-65.

Klar, A. J. (1998). "Propagating epigenetic states through meiosis: where Mendel's gene is more than a DNA moiety." Trends Genet **14**(8): 299-301.

Kobayashi, H., H. Hiura, et al. (2009). "DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm." Eur J Hum Genet **17**(12): 1582-91.

Kobayashi, H., A. Sato, et al. (2007). "Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients." Hum Mol Genet **16**(21): 2542-51.

Koch, C. M., C. V. Suschek, et al. (2011). "Specific age-associated DNA methylation changes in human dermal fibroblasts." PLoS One **6**(2): e16679.

Koch, C. M. and W. Wagner (2011). "Epigenetic-aging-signature to determine age in different tissues." Aging (Albany NY) **3**(10): 1018-27.

Kokkinaki, M., T. L. Lee, et al. (2010). "Age affects gene expression in mouse spermatogonial stem/progenitor cells." Reproduction **139**(6): 1011-20.

Kolevzon, A., R. Gross, et al. (2007). "Prenatal and perinatal risk factors for autism: a review and integration of findings." Arch Pediatr Adolesc Med **161**(4): 326-33.

Kondrashov, A. S. (2003). "Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases." Hum Mutat **21**(1): 12-27.

Konradi, C., M. Eaton, et al. (2004). "Molecular evidence for mitochondrial dysfunction in bipolar disorder." Arch Gen Psychiatry **61**(3): 300-8.

Krishnaswamy, S., K. Subramaniam, et al. (2011). "Delayed fathering and risk of mental disorders in adult offspring." Early Hum Dev **87**(3): 171-5.

Kuff, E. L. and K. K. Lueders (1988). "The intracisternal A-particle gene family: structure and functional aspects." Adv Cancer Res **51**: 183-276.

Kuiper, R. P., M. J. Ligtenberg, et al. (2010). "Germline copy number variation and cancer risk." Curr Opin Genet Dev.

Kumar, R. A., S. KaraMohamed, et al. (2008). "Recurrent 16p11.2 microdeletions in autism." Hum Mol Genet **17**(4): 628-38.

Kuratomi, G., K. Iwamoto, et al. (2008). "Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins." Mol Psychiatry **13**(4): 429-41.

Kuromitsu, J., A. Yokoi, et al. (2001). "Reduced neuropeptide Y mRNA levels in the frontal cortex of people with schizophrenia and bipolar disorder." Brain Res Gene Expr Patterns **1**(1): 17-21.

Kurukuti, S., V. K. Tiwari, et al. (2006). "CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2." Proc Natl Acad Sci U S A **103**(28): 10684-9.

Lachman, H. M., E. Pedrosa, et al. (2007). "Increase in GSK3beta gene copy number variation in bipolar disorder." Am J Med Genet B Neuropsychiatr Genet **144B**(3): 259-65.

Laird, P. W. (2010). "Principles and challenges of genomewide DNA methylation analysis." Nat Rev Genet **11**(3): 191-203.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lane, N., W. Dean, et al. (2003). "Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse." Genesis **35**(2): 88-93.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." BMC Bioinformatics **9**: 559.

Larsen, F., G. Gundersen, et al. (1992). "CpG islands as gene markers in the human genome." Genomics **13**(4): 1095-107.

Lauritsen, M. B., C. B. Pedersen, et al. (2005). "Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study." J Child Psychol Psychiatry **46**(9): 963-71.

Le Marechal, C., E. Masson, et al. (2006). "Hereditary pancreatitis caused by triplication of the trypsinogen locus." Nat Genet **38**(12): 1372-4.

Lee, C. K., R. Weindruch, et al. (2000). "Gene-expression profile of the ageing brain in mice." Nat Genet **25**(3): 294-7.

Lee, H., D. Malaspina, et al. (2011). "Paternal age related schizophrenia (PARS): Latent subgroups detected by k-means clustering analysis." Schizophr Res **128**(1-3): 143-9.

Lein, E. S., M. J. Hawrylycz, et al. (2007). "Genome-wide atlas of gene expression in the adult mouse brain." Nature **445**(7124): 168-76.

Leonhardt, H., A. W. Page, et al. (1992). "A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei." Cell **71**(5): 865-73.

Levinson, D. F., J. Duan, et al. (2011). "Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications." Am J Psychiatry **168**(3): 302-16.

Levitas, E., E. Lunenfeld, et al. (2007). "Relationship between age and semen parameters in men with normal sperm concentration: analysis of 6022 semen samples." Andrologia **39**(2): 45-50.

Li, E. (2002). "Chromatin modification and epigenetic reprogramming in mammalian development." Nat Rev Genet **3**(9): 662-73.

Li, J., T. Jiang, et al. (2004). "Genomic segmental polymorphisms in inbred mouse strains." Nat Genet **36**(9): 952-4.

Liang, G., M. F. Chan, et al. (2002). "Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements." Mol Cell Biol **22**(2): 480-91.

Lieber, M. R., Y. Ma, et al. (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol **4**(9): 712-20.

Lin, S. M., P. Du, et al. (2008). "Model-based variance-stabilizing transformation for Illumina microarray data." Nucleic Acids Res **36**(2): e11.

Lister, R., M. Pelizzola, et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-22.

Locke, D. P., A. J. Sharp, et al. (2006). "Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome." Am J Hum Genet **79**(2): 275-90.

Lopez-Castroman, J., D. D. Gomez, et al. (2010). "Differences in maternal and paternal age between schizophrenia and other psychiatric disorders." Schizophr Res **116**(2-3): 184-90.

Lubin, F. D., T. L. Roth, et al. (2008). "Epigenetic regulation of BDNF gene transcription in the consolidation of fear memory." J Neurosci **28**(42): 10576-86.

Luedi, P. P., A. J. Hartemink, et al. (2005). "Genome-wide prediction of imprinted murine genes." Genome Res **15**(6): 875-84.

Lundstrom, S., C. M. Haworth, et al. (2010). "Trajectories leading to autism spectrum disorders are affected by paternal age: findings from two nationally representative twin studies." J Child Psychol Psychiatry **51**(7): 850-6.

Ma, D. K., M. C. Marchetto, et al. (2010). "Epigenetic choreographers of neurogenesis in the adult mammalian brain." Nat Neurosci **13**(11): 1338-44.

Ma, D. Q., M. L. Cuccaro, et al. (2007). "Dissecting the locus heterogeneity of autism: significant linkage to chromosome 12q14." Mol Psychiatry **12**(4): 376-84.

Maher, E. R., L. A. Brueton, et al. (2003). "Beckwith-Wiedemann syndrome and assisted reproduction technology (ART)." J Med Genet **40**(1): 62-4.

Maimburg, R. D. and M. Vaeth (2006). "Perinatal risk factors and infantile autism." Acta Psychiatr Scand **114**(4): 257-64.

Malaspina, D., S. Harlap, et al. (2001). "Advancing paternal age and the risk of schizophrenia." Arch Gen Psychiatry **58**(4): 361-7.

Malaspina, D., A. Reichenberg, et al. (2005). "Paternal age and intelligence: implications for age-related genomic changes in male germ cells." Psychiatr Genet **15**(2): 117-25.

Mamtani, M., B. Rovin, et al. (2008). "CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus." Ann Rheum Dis **67**(8): 1076-83.

Mann, M. R. and M. S. Bartolomei (1999). "Towards a molecular understanding of Prader-Willi and Angelman syndromes." Hum Mol Genet **8**(10): 1867-73.

Marioni, J. C., N. P. Thorne, et al. (2007). "Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization." Genome Biol **8**(10): R228.

Marques, C. J., F. Carvalho, et al. (2004). "Genomic imprinting in disruptive spermatogenesis." Lancet **363**(9422): 1700-2.

Marshall, C. R., A. Noor, et al. (2008). "Structural variation of chromosomes in autism spectrum disorder." Am J Hum Genet **82**(2): 477-88.

Martin, P. and M. Albers (1995). "Cerebellum and schizophrenia: a selective review." Schizophr Bull **21**(2): 241-50.

Marutle, A., X. Zhang, et al. (2001). "Laminar distribution of nicotinic receptor subtypes in cortical regions in schizophrenia." J Chem Neuroanat **22**(1-2): 115-26.

Mattick, J. S. and I. V. Makunin (2006). "Non-coding RNA." Hum Mol Genet **15 Spec No 1**: R17-29.

Maziade, M., Y. C. Chagnon, et al. (2009). "Chromosome 13q13-q14 locus overlaps mood and psychotic disorders: the relevance for redefining phenotype." Eur J Hum Genet **17**(8): 1034-42.

McCarroll, S. A., T. N. Hadnott, et al. (2006). "Common deletion polymorphisms in the human genome." Nat Genet **38**(1): 86-92.

McCarroll, S. A., F. G. Kuruvilla, et al. (2008). "Integrated detection and population-genetic analysis of SNPs and copy number variation." Nat Genet **40**(10): 1166-74.

McCarthy, S. E., V. Makarov, et al. (2009). "Microduplications of 16p11.2 are associated with schizophrenia." Nat Genet **41**(11): 1223-7.

McIntosh, G. C., A. F. Olshan, et al. (1995). "Paternal age and the risk of birth defects in offspring." Epidemiology **6**(3): 282-8.

McKinney, C., M. E. Merriman, et al. (2008). "Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis." Ann Rheum Dis **67**(3): 409-13.

McMahon, F. J., O. C. Stine, et al. (1995). "Patterns of maternal transmission in bipolar affective disorder." Am J Hum Genet **56**(6): 1277-86.

McQuillin, A., N. Bass, et al. (2011). "Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample." Eur J Hum Genet **19**(5): 588-92.

Meaburn, E. L., L. C. Schalkwyk, et al. (2010). "Allele-specific methylation in the human genome: implications for genetic studies of complex disease." Epigenetics **5**(7): 578-82.

Medvedeva, Y. A., M. V. Fridman, et al. (2010). "Intergenic, gene terminal, and intragenic CpG islands in the human genome." BMC Genomics **11**: 48.

Menezes, P. R., G. Lewis, et al. (2010). "Paternal and maternal ages at conception and risk of bipolar affective disorder in their offspring." Psychol Med **40**(3): 477-85.

Migliore, L. and F. Coppede (2009). "Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases." Mutat Res **667**(1-2): 82-97.

Milanese, M., L. Segat, et al. (2009). "Copy number variation of defensin genes and HIV infection in Brazilian children." J Acquir Immune Defic Syndr **50**(3): 331-3.

Mill, J. (2007). "Rodent models: utility for candidate gene studies in human attention-deficit hyperactivity disorder (ADHD)." J Neurosci Methods **166**(2): 294-305.

Mill, J., T. Tang, et al. (2008). "Epigenomic profiling reveals DNA-methylation changes associated with major psychosis." Am J Hum Genet **82**(3): 696-711.

Millar, J. K., S. Mackie, et al. (2007). "Disrupted in schizophrenia 1 and phosphodiesterase 4B: towards an understanding of psychiatric illness." J Physiol **584**(Pt 2): 401-5.

Millar, J. K., B. S. Pickard, et al. (2005). "DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling." Science **310**(5751): 1187-91.

Millar, J. K., J. C. Wilson-Annan, et al. (2000). "Disruption of two novel genes by a translocation co-segregating with schizophrenia." Hum Mol Genet **9**(9): 1415-23.

Miller, B., E. Messias, et al. (2010). "Meta-analysis of Paternal Age and Schizophrenia Risk in Male Versus Female Offspring." Schizophr Bull.

Mirnics, K., F. A. Middleton, et al. (2001). "Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia." Mol Psychiatry **6**(3): 293-301.

Mizuguchi, T., R. Hashimoto, et al. (2008). "Microarray comparative genomic hybridization analysis of 59 patients with schizophrenia." J Hum Genet **53**(10): 914-9.

Moon, H. J., S. V. Yim, et al. (2006). "Identification of DNA copy-number aberrations by array-comparative genomic hybridization in patients with schizophrenia." Biochem Biophys Res Commun **344**(2): 531-9.

Moore, T. and D. Haig (1991). "Genomic imprinting in mammalian development: a parental tug-of-war." Trends Genet **7**(2): 45-9.

Moran, J. V., S. E. Holmes, et al. (1996). "High frequency retrotransposition in cultured mammalian cells." Cell **87**(5): 917-27.

Moreno-De-Luca, D., J. G. Mulle, et al. (2010). "Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia." Am J Hum Genet **87**(5): 618-30.

Morrow, E. M. (2010). "Genomic copy number variation in disorders of cognitive development." J Am Acad Child Adolesc Psychiatry **49**(11): 1091-104.

Morrow, E. M., S. Y. Yoo, et al. (2008). "Identifying autism loci and genes by tracing recent shared ancestry." Science **321**(5886): 218-23.

Mouridsen, S. E., B. Rich, et al. (1993). "Brief report: parental age in infantile autism, autistic-like conditions, and borderline childhood psychosis." J Autism Dev Disord **23**(2): 387-96.

Moy, S. S., J. J. Nadler, et al. (2004). "Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice." Genes Brain and Behavior **3**(5): 287-302.

Moy, S. S., J. J. Nadler, et al. (2004). "Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice." Genes Brain Behav **3**(5): 287-302.

Mueser, K. T., A. S. Bellack, et al. (1991). "Prevalence and stability of social skill deficits in schizophrenia." Schizophr Res **5**(2): 167-76.

Mulle, J. G., A. F. Dodd, et al. (2010). "Microdeletions of 3q29 confer high risk for schizophrenia." Am J Hum Genet **87**(2): 229-36.

Murgatroyd, C., A. V. Patchev, et al. (2009). "Dynamic DNA methylation programs persistent adverse effects of early-life stress." Nat Neurosci **12**(12): 1559-66.

Murphy, E. D. (1968). Characteristic Tumors. Biology of the Laboratory Mouse. E. L. Green. New York, DOVER PUBLICATIONS, INC.

Nagarajan, R. P., A. R. Hogart, et al. (2006). "Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation." Epigenetics **1**(4): e1-11.

Nakahata, Y., B. Grimaldi, et al. (2007). "Signaling to the circadian clock: plasticity by chromatin remodeling." Curr Opin Cell Biol **19**(2): 230-7.

Nakatani, N., E. Hattori, et al. (2006). "Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation." Hum Mol Genet **15**(12): 1949-62.

Naserbakht, M., H. R. Ahmadkhaniha, et al. (2011). "Advanced paternal age is a risk factor for schizophrenia in Iranians." Ann Gen Psychiatry **10**: 15.

Need, A. C., D. Ge, et al. (2009). "A genome-wide investigation of SNPs and CNVs in schizophrenia." PLoS Genet **5**(2): e1000373.

Nguyen, A., T. A. Rauch, et al. (2010). "Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism

candidate gene, RORA, whose protein product is reduced in autistic brain." FASEB J **24**(8): 3036-51.

Nguyen, D. Q., C. Webber, et al. (2006). "Bias of selection on human copy-number variants." PLoS Genet **2**(2): e20.

Nimblegen. (2011). "Comparative Genomic Hybridization (CGH)." from http://www.nimblegen.com/products/cgh/index.html.

Nolan, N. A. and M. W. Parkes (1973). "The effects of benzodiazepines on the behaviour of mice on a hole-board." Psychopharmacologia **29**(3): 277-86.

Notini, A. J., J. M. Craig, et al. (2008). "Copy number variation and mosaicism." Cytogenet Genome Res **123**(1-4): 270-7.

O'Donovan, M. C., G. Kirov, et al. (2008). "Phenotypic variations on the theme of CNVs." Nat Genet **40**(12): 1392-3.

Oakes, C. C., S. La Salle, et al. (2007). "A unique configuration of genome-wide DNA methylation patterns in the testis." Proc Natl Acad Sci U S A **104**(1): 228-33.

Ohara, K., H. D. Xu, et al. (1997). "Anticipation and imprinting in schizophrenia." Biol Psychiatry **42**(9): 760-6.

Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.

Ooi, S. K., C. Qiu, et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-7.

Orozco, L. D., S. J. Cokus, et al. (2009). "Copy number variation influences gene expression and metabolic traits in mice." Hum Mol Genet **18**(21): 4118-29.

Orstavik, K. H., K. Eiklid, et al. (2003). "Another case of imprinting defect in a girl with Angelman syndrome who was conceived by intracytoplasmic semen injection." Am J Hum Genet **72**(1): 218-9.

Padiath, Q. S., K. Saigoh, et al. (2006). "Lamin B1 duplications cause autosomal dominant leukodystrophy." Nat Genet **38**(10): 1114-23.

Paul, C., M. Nagano, et al. (2011). "Aging results in differential regulation of DNA repair pathways in pachytene spermatocytes in the brown norway rat." Biol Reprod **85**(6): 1269-78.

Paulsen, M. and A. C. Ferguson-Smith (2001). "DNA methylation in genomic imprinting, development, and disease." J Pathol **195**(1): 97-110.

Petersen, L., P. B. Mortensen, et al. (2010). "Paternal age at birth of first child and risk of schizophrenia." Am J Psychiatry **168**(1): 82-8.

Petronis, A., Gottesman, II, et al. (2003). "Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance?" Schizophr Bull **29**(1): 169-78.

Pickard, B. S., A. Christoforou, et al. (2009). "Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder." Mol Psychiatry **14**(9): 874-84.

Pickard, B. S., A. A. Pieper, et al. (2006). "The NPAS3 gene--emerging evidence for a role in psychiatric illness." Ann Med **38**(6): 439-48.

Pidsley, R., E. Dempster, et al. (2012). "Epigenetic and genetic variation at the IGF2/H19 imprinting control region on 11p15.5 is strongly associated with cerebellum weight." Epigenetics **7**(2).

Pidsley, R., E. L. Dempster, et al. (2010). "Brain weight in males is correlated with DNA methylation at IGF2." Mol Psychiatry **15**(9): 880-1.

Pidsley, R. and J. Mill (2011). "Epigenetic studies of psychosis: current findings, methodological approaches, and implications for postmortem research." Biol Psychiatry **69**(2): 146-56.

Pitman, S. D. (2001). "DNA Mutation Rates." from http://www.detectingdesign.com/dnamutationrates.html.

Piton, A., J. Gauthier, et al. (2011). "Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia." Mol Psychiatry **16**(8): 867-80.

Plagge, A., A. R. Isles, et al. (2005). "Imprinted Nesp55 influences behavioral reactivity to novel environments." Mol Cell Biol **25**(8): 3019-26.

Pongrac, J., F. A. Middleton, et al. (2002). "Gene expression profiling with DNA microarrays: advancing our understanding of psychiatric disorders." Neurochem Res **27**(10): 1049-63.

Popp, C., W. Dean, et al. (2010). "Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency." Nature **463**(7284): 1101-5.

Priebe, L., F. A. Degenhardt, et al. (2011). "Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder." Mol Psychiatry.

Pun, F. W., C. Zhao, et al. (2011). "Imprinting in the schizophrenia candidate gene GABRB2 encoding GABA(A) receptor beta(2) subunit." Mol Psychiatry **16**(5): 557-68.

Purcell, A. E., O. H. Jeon, et al. (2001). "Postmortem brain abnormalities of the glutamate neurotransmitter system in autism." Neurology **57**(9): 1618-28.

Qiagen. (2011). "DNeasy Blood & Tissue Kit." from http://www.qiagen.com/products/genomicdnastabilizationpurification/dneasytissuesystem/dneasybloodtissuekit.aspx#Tabs=t0.

Rakyan, V. and E. Whitelaw (2003). "Transgenerational epigenetic inheritance." Curr Biol **13**(1): R6.

Rakyan, V. K., M. E. Blewitt, et al. (2002). "Metastable epialleles in mammals." Trends Genet **18**(7): 348-51.

Rakyan, V. K., T. A. Down, et al. (2010). "Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains." Genome Res **20**(4): 434-9.

Ramocki, M. B., M. Bartnik, et al. (2010). "Recurrent distal 7q11.23 deletion including HIP1 and YWHAG identified in patients with intellectual disabilities, epilepsy, and neurobehavioral problems." Am J Hum Genet **87**(6): 857-65.

Redon, R., S. Ishikawa, et al. (2006). "Global variation in copy number in the human genome." Nature **444**(7118): 444-54.

Reichenberg, A., R. Gross, et al. (2006). "Advancing paternal age and autism." Arch Gen Psychiatry **63**(9): 1026-32.

Reichman, N. E. and J. O. Teitler (2006). "Paternal age as a risk factor for low birthweight." Am J Public Health **96**(5): 862-6.

Reik, W., W. Dean, et al. (2001). "Epigenetic reprogramming in mammalian development." Science **293**(5532): 1089-93.

Renthal, W. and E. J. Nestler (2009). "Histone acetylation in drug addiction." Semin Cell Dev Biol **20**(4): 387-94.

Retief, A. E., J. A. Van Zyl, et al. (1984). "Chromosome studies in 496 infertile males with a sperm count below 10 million/ml." Hum Genet **66**(2-3): 162-4.

Richards, E. J. (2006). "Inherited epigenetic variation--revisiting soft inheritance." Nat Rev Genet **7**(5): 395-401.

Riggs, A. D., R. A. Martienssen, et al. (1996). Introduction. Epigenetic Mechanisms of Gene Regulation V. E. A. Russo. New York, Cold Spring Harbor Laboratory Press. **32:** 1.

Riggs, A. D., Z. Xiong, et al. (1998). "Methylation dynamics, epigenetic fidelity and X chromosome structure." Novartis Found Symp **214**: 214-25; discussion 225-32.

Riley, B. P., A. Makoff, et al. (2000). "Haplotype transmission disequilibrium and evidence for linkage of the CHRNA7 gene region to schizophrenia in Southern African Bantu families." Am J Med Genet **96**(2): 196-201.

Ripke, S., A. R. Sanders, et al. (2011). "Genome-wide association study identifies five new schizophrenia loci." Nat Genet **43**(10): 969-76.

Robertson, K. D. and A. P. Wolffe (2000). "DNA methylation in health and disease." Nat Rev Genet **1**(1): 11-9.

Robins, L. N., J. E. Helzer, et al. (1984). "Lifetime prevalence of specific psychiatric disorders in three sites." Arch Gen Psychiatry **41**(10): 949-58.

Rodriguez-Santiago, B., A. Brunet, et al. (2010). "Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia." Mol Psychiatry **15**(10): 1023-33.

Roman-Gomez, J., A. Jimenez-Velasco, et al. (2005). "Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia." Oncogene **24**(48): 7213-23.

Roohi, J., C. Montagna, et al. (2009). "Disruption of contactin 4 in three subjects with autism spectrum disorder." J Med Genet **46**(3): 176-82.

Roohi, J., D. H. Tegay, et al. (2008). "A de novo apparently balanced translocation [46,XY,t(2;9)(p13;p24)] interrupting RAB11FIP5 identifies a potential candidate gene for autism spectrum disorder." Am J Med Genet B Neuropsychiatr Genet **147B**(4): 411-7.

Rosa, A., M. M. Picchioni, et al. (2008). "Differential methylation of the X-chromosome is a possible source of discordance for bipolar disorder female monozygotic twins." Am J Med Genet B Neuropsychiatr Genet **147B**(4): 459-62.

Roth, D. B. and J. H. Wilson (1986). "Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction." Mol Cell Biol **6**(12): 4295-304.

Rousseaux, S., C. Caron, et al. (2005). "Establishment of male-specific epigenetic information." Gene **345**(2): 139-53.

Ruder, A. (1985). "Paternal-age and birth-order effect on the human secondary sex ratio." Am J Hum Genet **37**(2): 362-72.

Rujescu, D., A. Ingason, et al. (2009). "Disruption of the neurexin 1 gene is associated with schizophrenia." Hum Mol Genet **18**(5): 988-96.

Saha, R. N. and K. Pahan (2006). "Regulation of inducible nitric oxide synthase gene in glial cells." Antioxid Redox Signal **8**(5-6): 929-47.

Saitou, M., S. C. Barton, et al. (2002). "A molecular programme for the specification of germ cell fate in mice." Nature **418**(6895): 293-300.

Salonia, A., R. Matloob, et al. (2011). "Are Caucasian-European men delaying fatherhood? Results of a 7 year observational study of infertile couples with male factor infertility." Int J Androl.

Samaco, R. C., A. Hogart, et al. (2005). "Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3." Hum Mol Genet **14**(4): 483-92.

Santos, F. and W. Dean (2004). "Epigenetic reprogramming during early development in mammals." Reproduction **127**(6): 643-51.

Sartorelli, E. M., L. F. Mazzucatto, et al. (2001). "Effect of paternal age on human sperm chromosomes." Fertil Steril **76**(6): 1119-23.

Sasaki, H. and Y. Matsui (2008). "Epigenetic events in mammalian germ-cell development: reprogramming and beyond." Nat Rev Genet **9**(2): 129-40.

Sasanfar, R., S. A. Haddad, et al. (2010). "Paternal age increases the risk for autism in an Iranian population sample." Mol Autism **1**(1): 2.

Sassaman, D. M., B. A. Dombroski, et al. (1997). "Many human L1 elements are capable of retrotransposition." Nat Genet **16**(1): 37-43.

Schalkwyk, L. C., E. L. Meaburn, et al. (2010). "Allelic skewing of DNA methylation is widespread across the genome." Am J Hum Genet **86**(2): 196-212.

Scherer, S. W., C. Lee, et al. (2007). "Challenges and standards in integrating surveys of structural variation." Nat Genet **39**(7 Suppl): S7-15.

Schmahmann, J. (2000). "The role of the cerebellum in affect and psychosis." Journal of Neurolinguistics **13**(2-3): 189-214.

Schroer, R. J., M. C. Phelan, et al. (1998). "Autism and maternally derived aberrations of chromosome 15q." Am J Med Genet **76**(4): 327-36.

Sebat, J., B. Lakshmi, et al. (2007). "Strong association of de novo copy number mutations with autism." Science **316**(5823): 445-9.

Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-8.

Sequenom. (2007). "Sequenom EpiTyper Beta." from http://www.epidesigner.com/.

Sequenom. (2011). "Sequenom EpiTyper." from http://www.sequenom.com/home/products---services/genetic-analysis/applications/epityper-dna-methylation-analysis/.

Serajee, F. J., H. Zhong, et al. (2003). "The metabotropic glutamate receptor 8 gene at 7q31: partial duplication and possible association with autism." J Med Genet **40**(4): e42.

Serretti, A. and L. Mandelli (2008). "The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions." Mol Psychiatry **13**(8): 742-71.

Sharp, A. J., D. P. Locke, et al. (2005). "Segmental duplications and copy-number variation in the human genome." Am J Hum Genet **77**(1): 78-88.

She, X., Z. Cheng, et al. (2008). "Mouse segmental duplication and copy number variation." Nat Genet **40**(7): 909-14.

Shelton, J. F., D. J. Tancredi, et al. (2010). "Independent and dependent contributions of advanced maternal and paternal ages to autism risk." Autism Res.

Shen, L., Y. Kondo, et al. (2007). "Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters." PLoS Genet **3**(10): 2023-36.

Shen, Y., J. Chow, et al. (2006). "Abnormal CpG island methylation occurs during in vitro differentiation of human embryonic stem cells." Hum Mol Genet **15**(17): 2623-35.

Siegmund, K. D., C. M. Connor, et al. (2007). "DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons." PLoS One **2**(9): e895.

Simon, N. M., M. W. Otto, et al. (2004). "Anxiety disorder comorbidity in bipolar disorder patients: data from the first 500 participants in the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD)." Am J Psychiatry **161**(12): 2222-9.

Sinclair, K. D., C. Allegrucci, et al. (2007). "DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status." Proc Natl Acad Sci U S A **104**(49): 19351-6.

Singer, M. F. (1982). "SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes." Cell **28**(3): 433-4.

Sipos, A., F. Rasmussen, et al. (2004). "Paternal age and schizophrenia: a population based cohort study." BMJ **329**(7474): 1070.

Skuse, D. H., R. S. James, et al. (1997). "Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function." Nature **387**(6634): 705-8.

Smith, E. N., C. S. Bloss, et al. (2009). "Genome-wide association study of bipolar disorder in European American and African American individuals." Mol Psychiatry **14**(8): 755-63.

Smith, R. and J. Mill (2011). Epigenetics and Chronic Diseases: An Overview. Epigenetic Aspects of Chronic Disease. H. I. Roach, F. Bronner and R. O. C. Oreffo. London, Springer-Verlag London Limited**:** 1-20.

Smith, R. G., R. L. Kember, et al. (2009). "Advancing paternal age is associated with deficits in social and exploratory behaviors in the offspring: a mouse model." PLoS One **4**(12): e8456.

Sparks, B. F., S. D. Friedman, et al. (2002). "Brain structural abnormalities in young children with autism spectrum disorder." Neurology **59**(2): 184-92.

Spiliotaki, M., V. Salpeas, et al. (2006). "Altered glucocorticoid receptor signaling cascade in lymphocytes of bipolar disorder patients." Psychoneuroendocrinology **31**(6): 748-60.

St Clair, D. (2009). "Copy number variation and schizophrenia." Schizophr Bull **35**(1): 9-12.

Statistics, O. o. N. (2009). Review of the National Statistician on births and patterns of family building in England and Wales, 2008. New Port, Office of National Statistics.

Stefansson, H., D. Rujescu, et al. (2008). "Large recurrent microdeletions associated with schizophrenia." Nature **455**(7210): 232-6.

Stone, J. L., M. O'Donovan, et al. (2008). "Rare chromosomal deletions and duplications increase risk of schizophrenia." Nature **455**(7210): 237-41.

Stranger, B. E., M. S. Forrest, et al. (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes." Science **315**(5813): 848-53.

Strom, S. P., J. L. Stone, et al. (2010). "High-density SNP association study of the 17q21 chromosomal region linked to autism identifies CACNA1G as a novel candidate gene." Mol Psychiatry **15**(10): 996-1005.

Stuss, D. T. and R. T. Knight (2002). Principles of Frontal Lobe Function. New York, Oxford University Press Inc.

Sun, X., J. F. Wang, et al. (2006). "Downregulation in components of the mitochondrial electron transport chain in the postmortem frontal cortex of subjects with bipolar disorder." J Psychiatry Neurosci **31**(3): 189-96.

Sun, Y., L. Zhang, et al. (2001). "Serial analysis of gene expression in the frontal cortex of patients with bipolar disorder." Br J Psychiatry Suppl **41**: s137-41.

Susser, E., R. Neugebauer, et al. (1996). "Schizophrenia after prenatal famine. Further evidence." Arch Gen Psychiatry **53**(1): 25-31.

Sykes, N. H., C. Toma, et al. (2009). "Copy number variation and association analysis of SHANK3 as a candidate gene for autism in the IMGSAC collection." Eur J Hum Genet **17**(10): 1347-53.

Takai, D. and P. A. Jones (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." Proc Natl Acad Sci U S A **99**(6): 3740-5.

Testart, J., E. Gautier, et al. (1996). "Intracytoplasmic sperm injection in infertile patients with structural chromosome abnormalities." Hum Reprod **11**(12): 2609-12.

Thanseem, I., A. Anitha, et al. (2011). "Elevated Transcription Factor Specificity Protein 1 in Autistic Brains Alters the Expression of Autism Candidate Genes." Biol Psychiatry.

Thomas, N. S., M. Durkie, et al. (2006). "Parental and chromosomal origin of unbalanced de novo structural chromosome abnormalities in man." Hum Genet **119**(4): 444-50.

Thomas, N. S., J. K. Morris, et al. (2010). "De novo apparently balanced translocations in man are predominantly paternal in origin and associated with a significant increase in paternal age." J Med Genet **47**(2): 112-5.

Tilghman, S. M. (1999). "The sins of the fathers and mothers: genomic imprinting in mammalian development." Cell **96**(2): 185-93.

Tkachev, D., M. L. Mimmack, et al. (2003). "Oligodendrocyte dysfunction in schizophrenia and bipolar disorder." Lancet **362**(9386): 798-805.

Torrey, E. F., B. M. Barci, et al. (2005). "Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains." Biol Psychiatry **57**(3): 252-60.

Torrey, E. F., S. Buka, et al. (2009). "Paternal age as a risk factor for schizophrenia: how important is it?" Schizophr Res **114**(1-3): 1-5.

Trasler, J. M. (1998). "Origin and roles of genomic methylation patterns in male germ cells." Semin Cell Dev Biol **9**(4): 467-74.

Tsuchiya, K. J., K. Matsumoto, et al. (2008). "Paternal age at birth and high-functioning autistic-spectrum disorder in offspring." Br J Psychiatry **193**(4): 316-21.

Tsuchiya, K. J., S. Takagai, et al. (2005). "Advanced paternal age associated with an elevated risk for schizophrenia in offspring in a Japanese population." Schizophr Res **76**(2-3): 337-42.

Tuzun, E., A. J. Sharp, et al. (2005). "Fine-scale structural variation of the human genome." Nat Genet **37**(7): 727-32.

Ullmann, R., G. Turner, et al. (2007). "Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation." Hum Mutat **28**(7): 674-82.

Ushijima, T., N. Watanabe, et al. (2003). "Fidelity of the methylation pattern and its variation in the genome." Genome Res **13**(5): 868-74.

Vacic, V., S. McCarthy, et al. (2011). "Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia." Nature **471**(7339): 499-503.

Vawter, M. P., H. Tomita, et al. (2006). "Mitochondrial-related gene expression changes are sensitive to agonal-pH state: implications for brain disorders." Mol Psychiatry **11**(7): 615, 663-79.

Voikar, V., A. Polus, et al. (2005). "Long-term individual housing in C57BL/6J and DBA/2 mice: assessment of behavioral consequences." Genes Brain Behav **4**(4): 240-52.

Voineagu, I., X. Wang, et al. (2011). "Transcriptomic analysis of autistic brain reveals convergent molecular pathology." Nature **474**(7351): 380-4.

Vorstman, J. A., W. G. Staal, et al. (2006). "Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism." Mol Psychiatry **11**(1): 1, 18-28.

Vrijenhoek, T., J. E. Buizer-Voskamp, et al. (2008). "Recurrent CNVs disrupt three candidate genes in schizophrenia patients." Am J Hum Genet **83**(4): 504-10.

Waddington, C. H. (1942). "The Epigenotype." Endeavour **1**(18).

Walsh, T., J. M. McClellan, et al. (2008). "Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia." Science **320**(5875): 539-43.

Walter, C. A., G. W. Intano, et al. (1998). "Mutation frequency declines during spermatogenesis in young mice but increases in old mice." Proc Natl Acad Sci U S A **95**(17): 10015-9.

Waterland, R. A. and R. L. Jirtle (2003). "Transposable elements: targets for early nutritional effects on epigenetic gene regulation." Mol Cell Biol **23**(15): 5293-300.

Watkins-Chow, D. E. and W. J. Pavan (2008). "Genomic copy number and expression variation within the C57BL/6J inbred mouse strain." Genome Res **18**(1): 60-6.

Weaver, I. C., N. Cervoni, et al. (2004). "Epigenetic programming by maternal behavior." Nat Neurosci **7**(8): 847-54.

Weaver, I. C., F. A. Champagne, et al. (2005). "Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life." J Neurosci **25**(47): 11045-54.

Weber, M. and D. Schubeler (2007). "Genomic patterns of DNA methylation: targets and function of an epigenetic mark." Curr Opin Cell Biol **19**(3): 273-80.

Weickert, C. S., R. E. Straub, et al. (2004). "Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain." Arch Gen Psychiatry **61**(6): 544-55.

Weiser, M., A. Reichenberg, et al. (2008). "Advanced parental age at birth is associated with poorer social functioning in adolescent males: shedding light on a core symptom of schizophrenia and autism." Schizophr Bull **34**(6): 1042-6.

Weiss, L. A., D. E. Arking, et al. (2009). "A genome-wide linkage and association scan reveals novel loci for autism." Nature **461**(7265): 802-8.

Weiss, L. A., Y. Shen, et al. (2008). "Association between microdeletion and microduplication at 16p11.2 and autism." N Engl J Med **358**(7): 667-75.

WellcomeTrustSangerInstitute. (2011). "The Copy Number Variation (CNV) Project."

Werner, S., D. Malaspina, et al. (2007). "Socioeconomic status at birth is associated with risk of schizophrenia: population-based multilevel study." Schizophr Bull **33**(6): 1373-8.

Whalley, L. J., B. M. Thomas, et al. (1995). "Epidemiology of presenile Alzheimer's disease in Scotland (1974-88) II. Exposures to possible risk factors." Br J Psychiatry **167**(6): 732-8.

Wikipedia. (2011). "Combined bisulfite restriction analysis." from http://en.wikipedia.org/wiki/File:Bisulfite_conversion.svg.

Wilkins, J. F. and D. Haig (2003). "What good is genomic imprinting: the function of parent-specific gene expression." Nat Rev Genet **4**(5): 359-68.

Wilkinson, L. S., W. Davies, et al. (2007). "Genomic imprinting effects on brain development and function." Nat Rev Neurosci **8**(11): 832-43.

Wilson, G. M., S. Flibotte, et al. (2006). "DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling." Hum Mol Genet **15**(5): 743-9.

Wilson, V. L. and P. A. Jones (1983). "DNA methylation decreases in aging but not in immortal cells." Science **220**(4601): 1055-7.

Winslow, J. T. (2003). "Mouse social recognition and preference." Curr Protoc Neurosci **Chapter 8**: Unit 8 16.

Wong, C. C., A. Caspi, et al. (2010). "A longitudinal study of epigenetic variation in twins." Epigenetics **5**(6): 516-26.

Wyrobek, A. J., B. Eskenazi, et al. (2006). "Advancing age has differential effects on DNA damage, chromatin integrity, gene mutations, and aneuploidies in sperm." Proc Natl Acad Sci U S A **103**(25): 9601-6.

Xiang, N., R. Zhao, et al. (2008). "Selenite reactivates silenced genes by modifying DNA methylation and histones in prostate cancer cells." Carcinogenesis **29**(11): 2175-81.

Xu, B., J. L. Roos, et al. (2008). "Strong association of de novo copy number mutations with sporadic schizophrenia." Nat Genet **40**(7): 880-5.

Xu, B., A. Woodroffe, et al. (2009). "Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans." Proc Natl Acad Sci U S A **106**(39): 16746-51.

Yalcin, B., K. Wong, et al. (2011). "Sequence-based characterization of structural variation in the mouse genome." Nature **477**(7364): 326-9.

Yang, S., T. Liu, et al. (2008). "Comparative proteomic analysis of brains of naturally aging mice." Neuroscience **154**(3): 1107-20.

Yang, Y., E. K. Chung, et al. (2007). "Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans." Am J Hum Genet **80**(6): 1037-54.

Yauk, C., A. Polyzos, et al. (2008). "Germ-line mutations, DNA damage, and global hypermethylation in mice exposed to particulate air pollution in an urban/industrial location." Proc Natl Acad Sci U S A **105**(2): 605-10.

Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." Trends Genet **13**(8): 335-40.

Zammit, S., P. Allebeck, et al. (2003). "Paternal age and risk for schizophrenia." Br J Psychiatry **183**: 405-8.

Zeier, Z., I. Madorsky, et al. (2011). "Gene expression in the hippocampus: regionally specific effects of aging and caloric restriction." Mech Ageing Dev **132**(1-2): 8-19.

Zhang, D., L. Cheng, et al. (2009). "Singleton deletions throughout the genome increase risk of bipolar disorder." Mol Psychiatry **14**(4): 376-80.

Zhang, H., X. Liu, et al. (2002). "Reelin gene alleles and susceptibility to autism spectrum disorders." Mol Psychiatry **7**(9): 1012-7.

Zhang, S. L. (1992). "[A study on effects of parents age, birth order and mental retardation of unknown etiology]." Zhonghua Shen Jing Jing Shen Ke Za Zhi **25**(5): 303-5, 318.

Zhang, X., C. C. Lv, et al. (2010). "Prenatal and perinatal risk factors for autism in China." J Autism Dev Disord **40**(11): 1311-21.

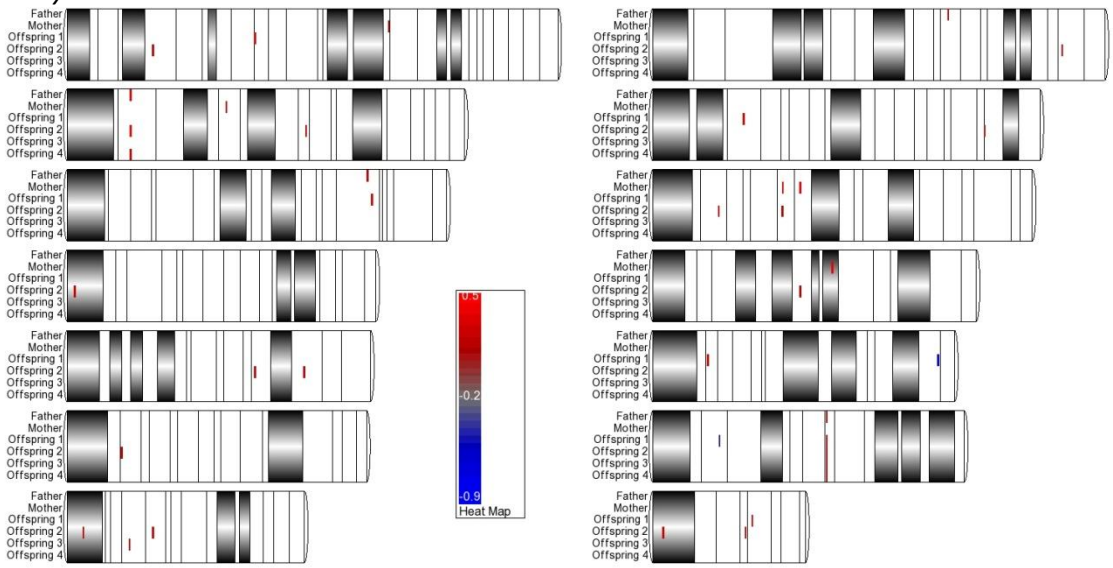
ZymoResearch. (2011). "EZ DNA Methylation Kit." from http://www.zymoresearch.com/zrc/pdf/D5001i.pdf.

Appendix

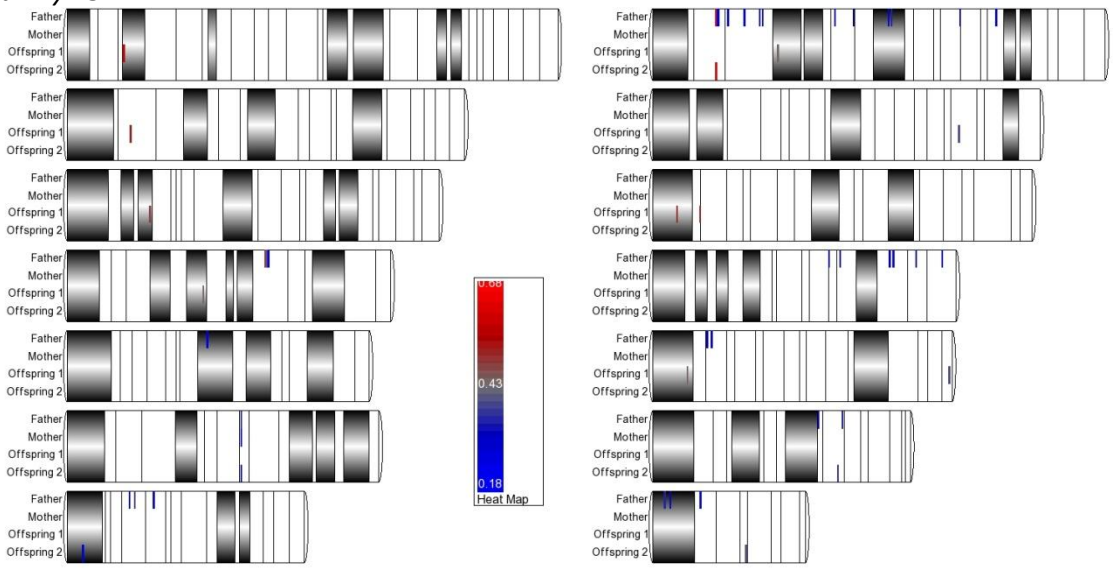
Appendix 1 - All CNVs by Family

Young Fathers

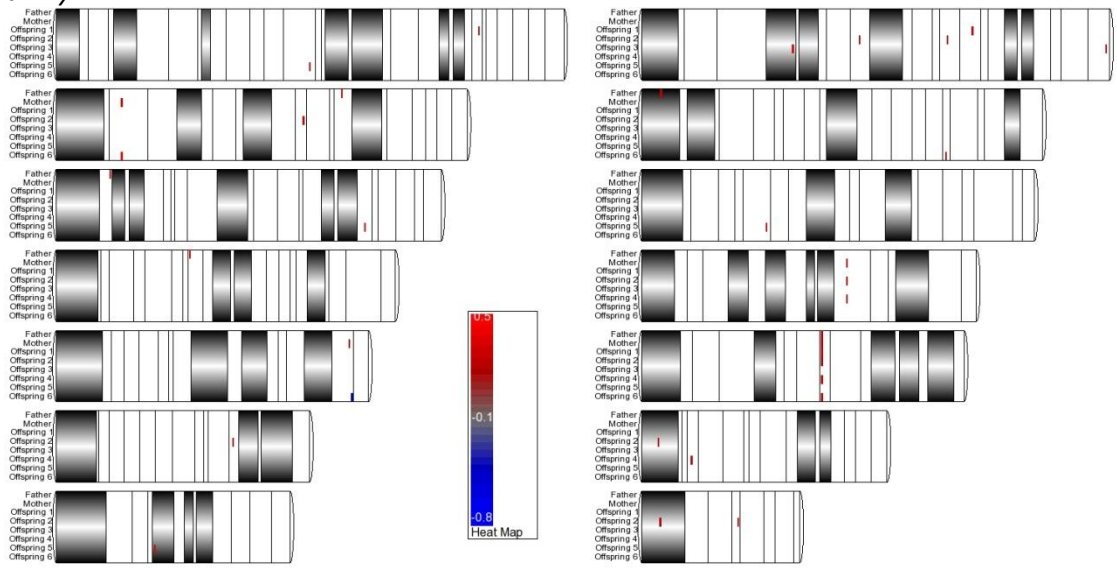
Family 11



Family 13

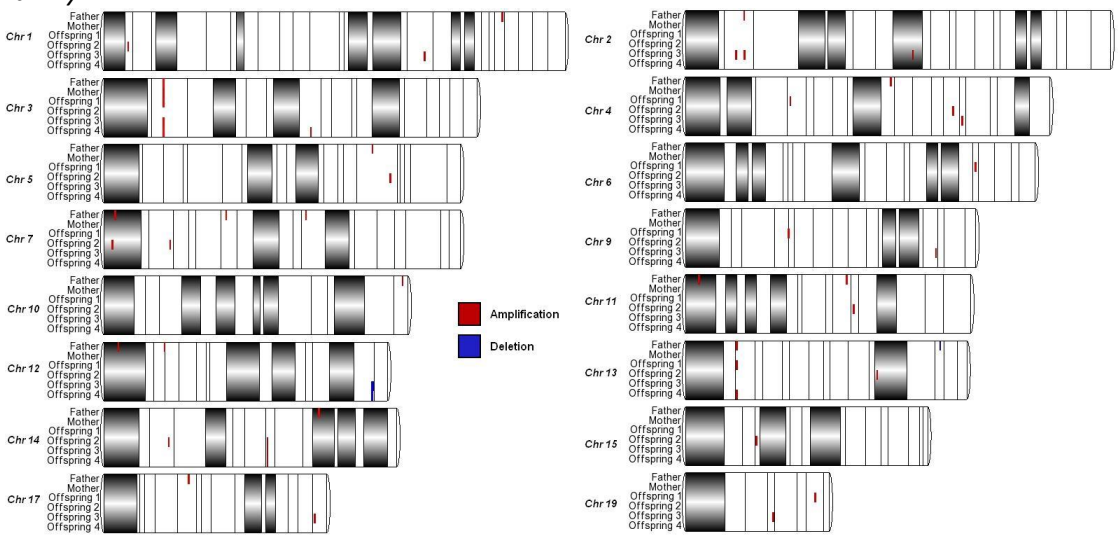


Family 14

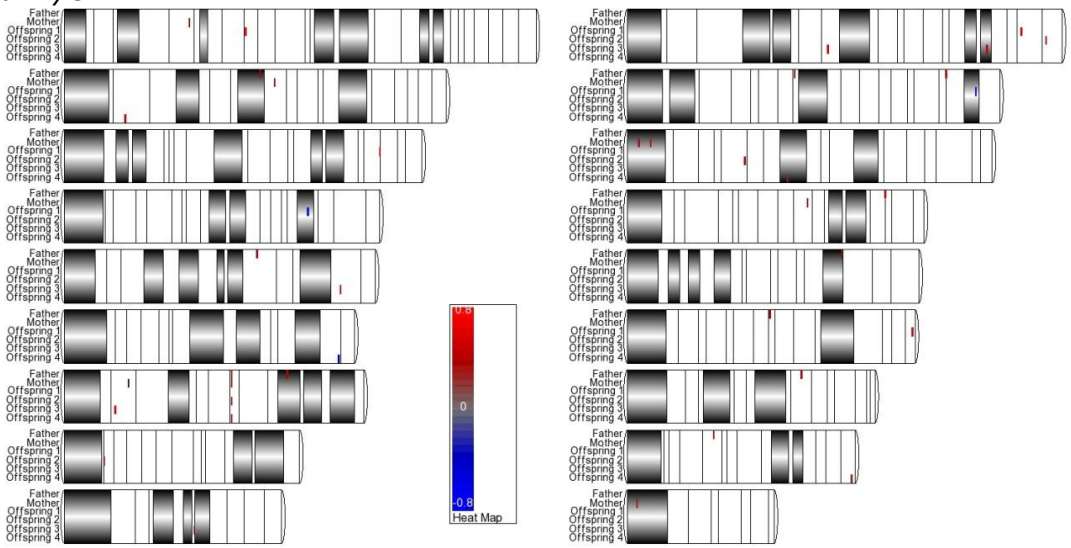


Old Fathers

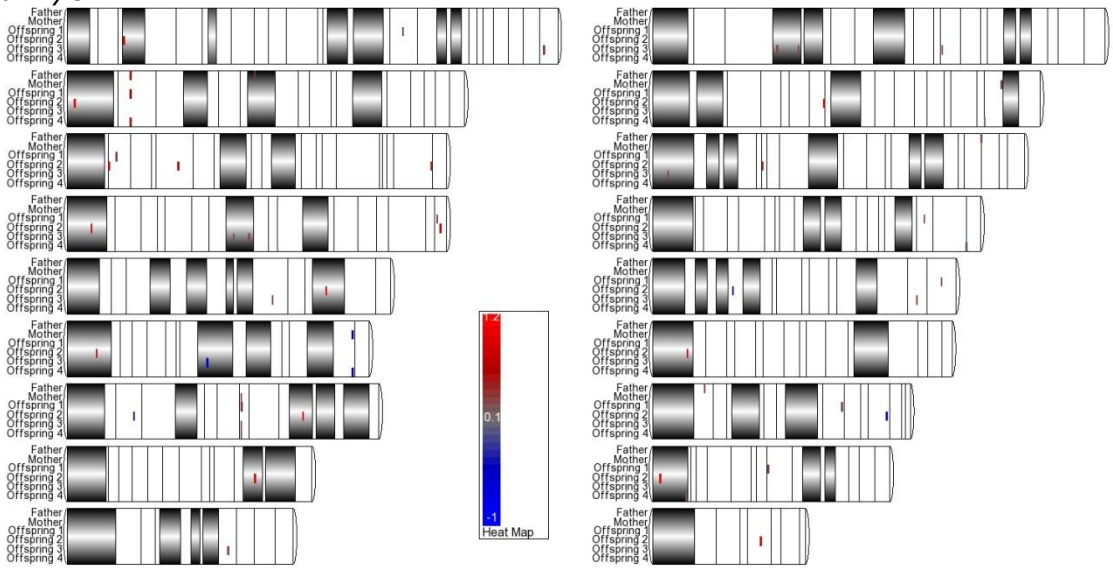
Family 2



Family 3

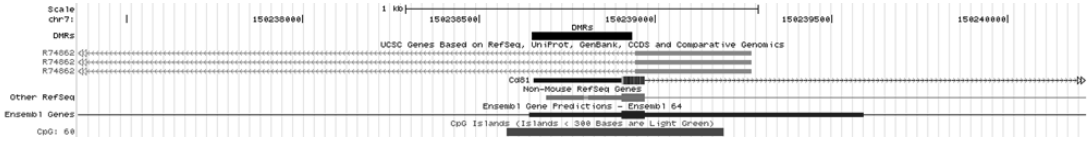


Family 5

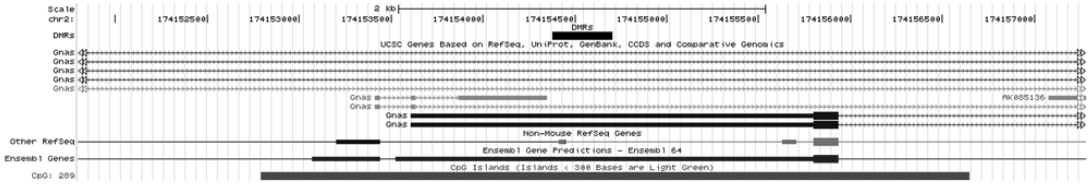

Father
Mother
Offspring 1
Offspring 2
Offspring 3
Offspring 4
1
0.1
-1
Heat Map

Appendix 2 - Chromosomal Location of DMRs

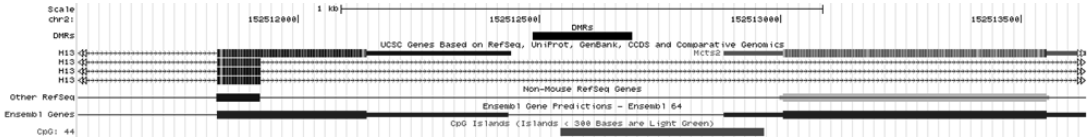
Cd81



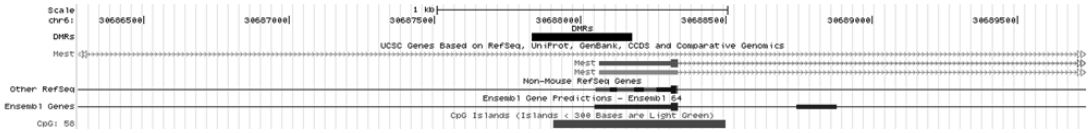
Exon1A


Scale
chr2:
174152500
174153000
174153500
174154000
174154500
174155000
174155500
174156000
174156500
174157000
2 kb
DMRs
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
Gnas
Gnas
Gnas

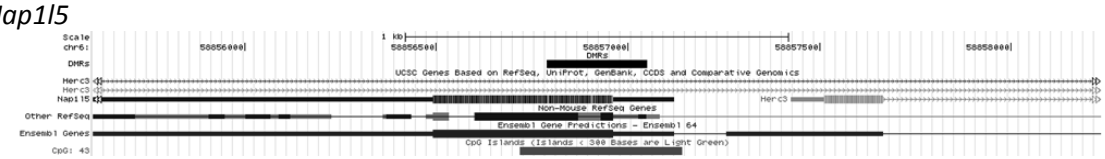
Mcts2



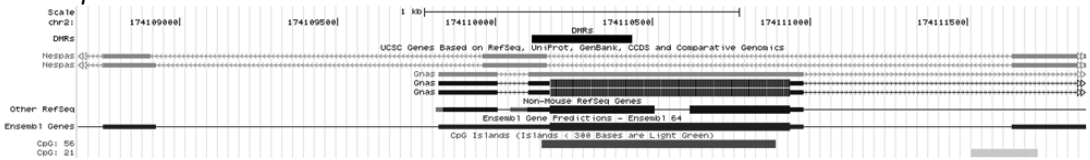
Mest



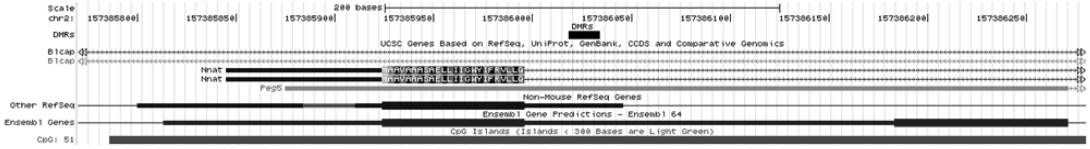
Nap1l5



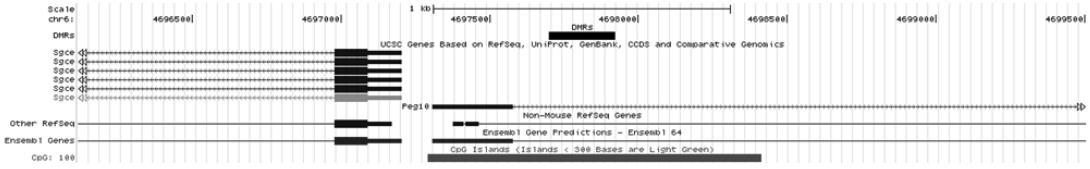
Gnas-Nesp



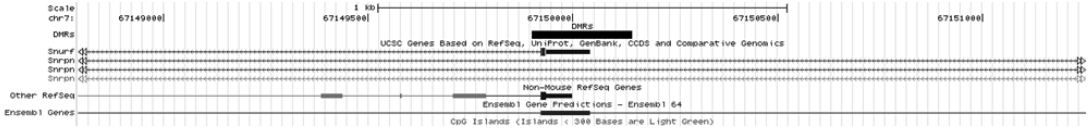
Nnat



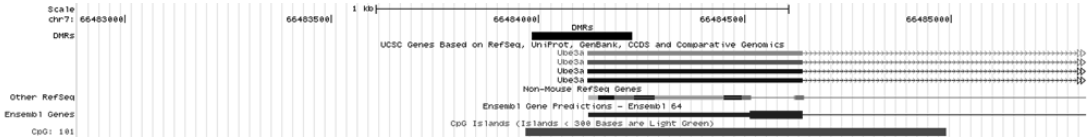
Sgce



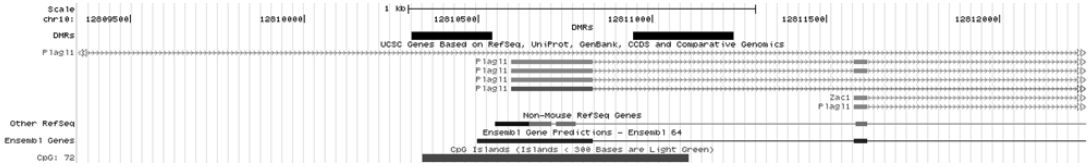
Snrpn



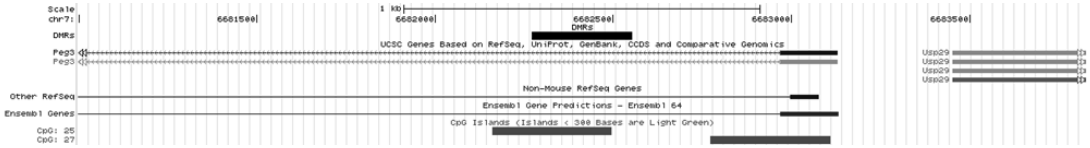
Ube3a



Zac1



Zim2



Appendix 3 – R Scripts

lumi

Load in Data

```
Expression <- 'ExpressionRGenesNames.txt'

example.lumi <- lumiR.batch(Expression)
```

Variance Stabilizing Transform

```
lumi.T <- lumiT(example.lumi)
```

RSN between Microarray Normaliazation

```
lumi.N<-lumiN(lumi.T, method= 'rsn')
```

Quality Control Estimation after Normalization

```
lumi.N.Q <- lumiQ(lumi.N)
```

Assign Group Identification to Data and by Group Analysis

```
sampleType <- c('Y', 'O', 'Y', 'O', 'Y', 'O', 'Y', 'O', 'Y', 'O', 'Y', 'O', 'Y', 'Y', 'Y', 'O', 'Y', 'O', 'Y', 'O','Y', 'Y',  'O','Y', 'O','Y', 'O','Y', 'O', 'O', 'O', 'O')

design <- model.matrix(~ factor(sampleType))

    colnames(design) <- c('Y', 'O-Y')

    fit <- lmFit(selDataMatrix, design)

    fit <- eBayes(fit)
```

WGCNA

Checking data for excessive missing values and identification of outlier microarray samples

```
gsg = goodSamplesGenes(datExpr0, verbose = 3);
gsg$allOK
```

Choose a set of soft-thresholding powers

```
powers = c(c(1:10), seq(from = 12, to=20, by=2))
```

Call the network topology analysis function

```
sft = pickSoftThreshold(datExpr0, powerVector = powers, verbose = 5, networkType = "signed")
```

Constructing the gene network and identifying modules

```
net100 = blockwiseModules(datExpr0, power = 10, minModuleSize = 100,
reassignThreshold = 0, mergeCutHeight = 0.25,
numericLabels = TRUE, pamRespectsDendro = FALSE,
saveTOMs = TRUE,
saveTOMFileBase = "Expr100signed",
verbose = 3, networkType = "signed")
table(net100$colors)
```

Recalculate MEs with colour labels

```
nGenes = ncol(datExpr0);

nSamples = nrow(datExpr0);

MEs0 = moduleEigengenes(datExpr0, moduleColors)$eigengenes

MEs = orderMEs(MEs0)

moduleTraitCor = cor(MEs, datTraits, use = "p");

moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples);
```

Appendix 4 – *Log2 Expression Changes Associated with CNVs*

Table showing log2 expression values between individuals containing CNVs in the listed genes and those without for genes ranked by fold change. 'Average' is the average expression across all individuals regardless of CNV status.

Gene	Individuals with CNV	Average CNV	Average	Fold Change	t-value	p-value	Direction of Change	CNV Type
PITPNM1	1	7.10	7.46	1.28	6.73	1.57E-07	↓	Amplification
EDF1	1	9.69	10.00	1.24	-2.10	0.044	↓	Amplification
IFIT3	1	7.89	7.59	1.23	1.93	0.063	↑	Amplification
CORO1B	1	7.61	7.85	1.18	-8.10	3.82E-09	↓	Amplification
CHPF	1	6.49	6.73	1.18	4.85	3.26E-05	↓	Amplification
D14ERTD449E	2	7.33	7.56	1.17	1.94	0.061	↓	Deletion
HBA-A1	1	12.80	13.00	1.15	-6.08	9.74E-07	↓	Deletion
CRK	1	8.81	8.62	1.14	3.97	3.93E-04	↑	Amplification
VEGFB	1	10.36	10.54	1.13	-4.27	1.73E-04	↓	Amplification
HIST1H2AO	3	10.90	10.72	1.13	-4.01	3.60E-04	↑	Amplification
ANXA11	2	7.63	7.80	1.13	-2.25	0.031	↓	Deletion
RPS6KB2	1	6.80	6.63	1.13	6.85	1.11E-07	↑	Amplification
TMEM198	1	8.63	8.46	1.12	-7.09	5.85E-08	↑	Amplification
TBC1D5	1	7.46	7.61	1.11	7.88	6.81E-09	↓	Amplification
PITPNM3	1	8.20	8.35	1.11	-4.37	1.29E-04	↓	Amplification
DBNDD1	1	7.34	7.19	1.11	7.37	2.66E-08	↑	Amplification
UBE3B	1	9.06	8.91	1.11	14.72	1.56E-15	↑	Amplification
EFNA2	1	6.65	6.50	1.11	0.00	1.000	↑	Amplification
CDK2AP2	1	8.12	7.98	1.10	0.05	0.963	↑	Amplification
MACROD1	1	7.26	7.40	1.10	2.72	0.011	↓	Amplification
IL4I1	1	6.65	6.78	1.10	-1.72	0.095	↓	Amplification
A630095E13RIK	1	6.47	6.60	1.10	5.35	7.86E-06	↓	Amplification
IMP4	1	8.12	7.99	1.09	0.09	0.930	↑	Amplification
BOLA1	1	7.95	7.82	1.09	-0.12	0.903	↑	Amplification
VTI1A	1	7.26	7.39	1.09	4.18	2.19E-04	↓	Amplification
TMEM134	1	8.11	7.99	1.09	-0.13	0.894	↑	Amplification
POLD4	1	7.76	7.64	1.09	-0.43	0.674	↑	Amplification
IFIT2	1	6.98	6.86	1.09	0.45	0.653	↑	Amplification
SPNB4	1	7.08	6.96	1.09	-0.61	0.545	↑	Amplification
YIPF3	1	8.39	8.51	1.09	-3.58	0.001	↓	Amplification
TNNC2	1	6.71	6.59	1.09	0.70	0.490	↑	Amplification
HIST2H3C1	2	7.37	7.25	1.09	-0.85	0.401	↑	Amplification
MED19	1	7.58	7.70	1.09	11.81	5.19E-13	↓	Amplification
RHOU	1	7.84	7.72	1.09	1.08	0.287	↑	Amplification
OLFR19	1	6.53	6.65	1.08	-11.53	9.66E-13	↓	Amplification
TJAP1	1	6.82	6.93	1.08	-9.17	2.46E-10	↓	Amplification

Gene	Individuals with CNV	Average CNV	Average	Fold Change	t-value	p-value	Direction of Change	CNV Type
IGFALS	1	6.73	6.62	1.08	-1.26	0.216	↑	Amplification
2310003C23RIK	1	8.75	8.64	1.08	1.36	0.182	↑	Amplification
TMEM101	1	7.70	7.59	1.08	-1.54	0.134	↑	Amplification
IVD	1	8.61	8.72	1.08	9.14	2.60E-10	↓	Amplification
DNM3OS	1	6.69	6.59	1.08	-1.63	0.114	↑	Amplification
FBF1	1	7.09	7.19	1.07	-5.48	5.43E-06	↓	Amplification
NUP62	1	7.66	7.56	1.07	1.77	0.087	↑	Amplification
ACOT8	1	7.92	7.82	1.07	1.84	0.076	↑	Amplification
CETN3	1	9.80	9.90	1.07	2.05	0.049	↓	Amplification
TFPT	1	7.32	7.42	1.07	3.58	0.001	↓	Amplification
HIST2H2AC	2	10.51	10.42	1.07	2.17	0.038	↑	Amplification
D830046C22RIK	1	6.44	6.53	1.06	8.43	1.60E-09	↓	Amplification
GAS8	1	6.97	7.06	1.06	3.27	0.003	↓	Amplification
SYT11	1	7.59	7.68	1.06	-2.89	0.007	↓	Amplification
MRPS34	1	11.03	10.94	1.06	-2.31	0.028	↑	Amplification
DNAJC4	1	7.21	7.30	1.06	2.51	0.018	↓	Amplification
BAHD1	1	7.58	7.67	1.06	-1.36	0.185	↓	Amplification
FKBP2	1	9.46	9.37	1.06	2.34	0.026	↑	Amplification
AIP	1	11.12	11.21	1.06	11.79	5.41E-13	↓	Amplification
2810403A07RIK	1	8.49	8.40	1.06	-2.49	0.018	↑	Amplification
AP2A1	1	8.69	8.78	1.06	2.01	0.053	↓	Amplification
PRPF31	1	6.60	6.68	1.06	1.17	0.250	↓	Amplification
CYP8B1	1	6.57	6.65	1.06	1.08	0.290	↓	Amplification
FERMT3	1	6.87	6.79	1.06	-2.68	0.012	↑	Amplification
DNM3	1	6.42	6.50	1.06	1.06	0.298	↓	Amplification
TULP4	1	9.44	9.36	1.06	-2.71	0.011	↑	Amplification
RIT1	1	9.28	9.36	1.06	-0.75	0.460	↓	Amplification
CNGA4	1	6.73	6.65	1.06	-2.85	0.008	↑	Amplification
STIP1	1	10.19	10.11	1.06	3.02	0.005	↑	Amplification
TRIM65	1	6.73	6.65	1.06	3.15	0.004	↑	Amplification
PPP1CA	1	11.12	11.04	1.05	-3.41	0.002	↑	Amplification
PHPT1	1	9.67	9.74	1.05	-0.31	0.756	↓	Amplification
HIST2H2AA1	2	7.58	7.50	1.05	3.97	3.92E-04	↑	Amplification
GPR146	1	6.83	6.90	1.05	-7.72	1.04E-08	↓	Amplification
MMP13	1	6.56	6.63	1.05	6.92	9.27E-08	↓	Amplification
MAFK	1	6.49	6.56	1.05	6.77	1.41E-07	↓	Amplification
SERTAD3	1	7.00	6.93	1.05	4.15	2.40E-04	↑	Amplification
HIST2H2AB	2	6.81	6.74	1.05	4.25	1.81E-04	↑	Amplification
CCBP2	1	7.01	6.94	1.05	-4.38	1.27E-04	↑	Amplification
FGFBP3	1	6.99	6.92	1.05	-4.38	1.25E-04	↑	Amplification
C130050O18RIK	1	6.68	6.61	1.05	4.44	1.06E-04	↑	Amplification
TMEM125	3	7.69	7.62	1.05	-4.75	4.40E-05	↑	Amplification
MYO1C	1	6.73	6.66	1.05	-5.38	7.15E-06	↑	Amplification

Gene	Individuals with CNV	Average CNV	Average	Fold Change	t-value	p-value	Direction of Change	CNV Type
HTT	1	6.77	6.71	1.05	-5.48	5.41E-06	↑	Amplification
BLVRB	1	8.80	8.87	1.05	6.19	7.14E-07	↓	Amplification
HAGH	1	10.13	10.07	1.05	-6.02	1.17E-06	↑	Amplification
RPS6KA1	2	7.64	7.70	1.04	-6.05	1.08E-06	↓	Amplification
MMP9	1	6.66	6.72	1.04	6.03	1.14E-06	↓	Amplification
MICALL2	1	6.60	6.66	1.04	4.95	2.50E-05	↓	Amplification
NUBP2	1	7.66	7.60	1.04	-6.33	4.75E-07	↑	Amplification
SERTAD1	1	7.08	7.02	1.04	-6.45	3.40E-07	↑	Amplification
APPL1	1	8.06	8.00	1.04	-6.82	1.23E-07	↑	Amplification
CNOT3	1	6.92	6.86	1.04	-7.77	9.17E-09	↑	Amplification
RAD9	1	6.70	6.76	1.04	-4.37	1.28E-04	↓	Amplification
INTS1	1	9.04	9.09	1.04	-3.91	4.66E-04	↓	Amplification
HIST2H3B	1	7.71	7.66	1.04	-7.78	8.81E-09	↑	Amplification
3110082I17RIK	1	6.96	7.01	1.04	3.87	5.23E-04	↓	Amplification
HIST2H2BB	1	6.57	6.62	1.04	-3.64	9.84E-04	↓	Amplification
ZFAND2A	1	9.71	9.76	1.04	-3.55	0.001	↓	Amplification
CDH22	1	6.74	6.79	1.03	2.73	0.010	↓	Amplification
UNCX	1	6.62	6.57	1.03	8.04	4.49E-09	↑	Amplification
CPHX	2	6.42	6.47	1.03	-2.67	0.012	↓	Deletion
ATP5D	1	9.96	9.91	1.03	-8.52	1.25E-09	↑	Amplification
TCERG1L	1	6.63	6.58	1.03	-8.94	4.32E-10	↑	Amplification
ACOX1	1	6.50	6.54	1.03	2.59	0.015	↓	Amplification
ATM	1	6.62	6.66	1.03	2.58	0.015	↓	Amplification
TBC1D17	1	8.39	8.35	1.03	-9.09	2.98E-10	↑	Amplification
NAGS	1	6.49	6.53	1.03	1.47	0.153	↓	Amplification
PTPN18	1	6.81	6.77	1.03	9.25	1.98E-10	↑	Amplification
ZCCHC7	1	6.67	6.63	1.03	-9.48	1.14E-10	↑	Amplification
MMAB	1	6.90	6.94	1.03	1.17	0.251	↓	Amplification
CABP4	1	6.53	6.57	1.03	-0.39	0.699	↓	Amplification
ZSWIM3	1	7.19	7.23	1.03	-0.17	0.869	↓	Amplification
COX19	1	9.30	9.34	1.03	-8.17	3.15E-09	↓	Amplification
HIST2H2BE	2	6.84	6.80	1.03	-11.29	1.64E-12	↑	Amplification
TGFB1	1	6.69	6.73	1.03	-5.25	1.04E-05	↓	Amplification
STK11	1	8.10	8.06	1.03	12.08	2.92E-13	↑	Amplification
PTPRCAP	1	6.45	6.49	1.03	3.91	4.65E-04	↓	Amplification
TRPT1	1	7.88	7.84	1.03	-14.52	2.26E-15	↑	Amplification
OLFR199	1	6.62	6.58	1.03	14.58	2.03E-15	↑	Amplification
CBL	1	6.63	6.59	1.03	-2.91	0.007	↑	Amplification
AKT1S1	1	8.46	8.50	1.03	-0.62	0.539	↓	Amplification
DOS	1	9.39	9.35	1.03	-3.87	5.31E-04	↑	Amplification
CUX2	2	7.22	7.26	1.03	6.44	3.55E-07	↓	Amplification
BICD1	1	6.63	6.66	1.02	2.17	0.038	↓	Amplification
DUXBL	2	6.62	6.59	1.02	15.05	8.42E-16	↑	Deletion

Gene	Individuals with CNV	Average CNV	Average	Fold Change	t-value	p-value	Direction of Change	CNV Type
1600002K03RIK	2	7.63	7.66	1.02	12.48	1.26E-13	↓	Amplification
EXOC4	1	7.94	7.91	1.02	-2.67	0.012	↑	Amplification
MYST2	1	7.40	7.37	1.02	0.69	0.495	↑	Amplification
HIST1H2AN	2	7.70	7.67	1.02	-3.72	7.81E-04	↑	Amplification
MRPL38	1	9.91	9.94	1.02	8.47	1.43E-09	↓	Amplification
TMEM184A	1	6.54	6.51	1.02	-3.73	7.79E-04	↑	Amplification
CPT1C	1	7.24	7.27	1.02	-6.22	6.57E-07	↓	Amplification
ACCN4	1	6.56	6.59	1.02	6.21	6.81E-07	↓	Amplification
MRGPRD	1	6.57	6.60	1.02	4.23	1.92E-04	↓	Amplification
ZFP335	1	7.04	7.01	1.02	-9.55	9.62E-11	↑	Amplification
ARID1A	3	6.60	6.58	1.02	-1.38	0.179	↑	Amplification
ESR2	1	6.58	6.61	1.02	1.80	0.081	↓	Amplification
CCDC97	1	6.71	6.68	1.02	-2.91	0.007	↑	Amplification
CIRBP	2	8.67	8.69	1.02	-1.17	0.253	↓	Amplification
OLFR692	1	6.66	6.64	1.02	-3.07	0.004	↑	Amplification
AFG3L1	1	7.27	7.29	1.02	-0.42	0.676	↓	Amplification
GPR152	1	6.59	6.57	1.02	3.14	0.004	↑	Amplification
SNX21	1	8.42	8.44	1.02	-5.86	1.84E-06	↓	Amplification
BAI1	1	6.86	6.84	1.02	3.91	4.71E-04	↑	Amplification
PNKP	1	8.13	8.15	1.01	-2.29	0.029	↓	Amplification
ADAMTS20	1	6.58	6.60	1.01	-2.01	0.054	↓	Deletion
NUDT22	1	7.62	7.60	1.01	-5.58	4.09E-06	↑	Amplification
FGF14	1	6.63	6.61	1.01	6.33	4.87E-07	↑	Amplification
SIGLEC5	1	6.57	6.55	1.01	-6.95	8.57E-08	↑	Amplification
TPCN2	1	6.69	6.71	1.01	-1.19	0.242	↓	Amplification
LENG1	1	6.71	6.70	1.01	8.25	2.56E-09	↑	Amplification
CDH1	1	6.75	6.74	1.01	9.13	2.70E-10	↑	Amplification
MAMDC4	1	6.59	6.60	1.01	-1.05	0.301	↓	Amplification
ACCN1	1	7.29	7.30	1.01	-15.80	2.20E-16	↓	Amplification
SPSB3	1	9.01	9.02	1.01	9.01	3.67E-10	↓	Amplification
SCAF1	1	8.90	8.89	1.01	-9.72	6.35E-11	↑	Amplification
SLC12A5	1	6.55	6.54	1.01	-1.19	0.241	↑	Amplification
GRK4	1	6.55	6.56	1.01	8.29	2.32E-09	↓	Amplification
PSG28	1	6.62	6.61	1.01	-2.28	0.030	↑	Amplification
MIDN	2	8.62	8.61	1.01	-1.27	0.213	↑	Amplification
MRGPRF	1	6.78	6.77	1.01	-2.44	0.020	↑	Amplification
ANAPC1	1	6.54	6.53	1.01	5.52	4.79E-06	↑	Amplification
H2-EB1	1	6.88	6.87	1.01	-1.13	0.269	↑	Amplification
IRF3	1	7.54	7.55	1.01	6.96	8.37E-08	↓	Amplification
KIF11	3	6.61	6.60	1.01	7.32	3.09E-08	↑	Amplification
DIDO1	1	6.83	6.82	1.01	-8.57	1.12E-09	↑	Amplification
SFI1	2	6.75	6.75	1.01	-5.80	2.14E-06	↓	Amplification
BCL2L12	1	6.71	6.72	1.01	5.60	3.80E-06	↓	Amplification

Gene	Individuals with CNV	Average CNV	Average	Fold Change	t-value	p-value	Direction of Change	CNV Type
ADAM22	1	6.93	6.94	1.01	5.54	4.62E-06	↓	Amplification
TBC1D10C	1	6.73	6.72	1.01	8.20	2.92E-09	↑	Amplification
HIST1H4M	3	6.66	6.67	1.00	4.81	3.73E-05	↓	Amplification
HIST2H4	1	6.60	6.59	1.00	-0.43	0.671	↑	Amplification
TSPAN4	1	6.55	6.54	1.00	-0.47	0.644	↑	Amplification
SPACA3	1	6.51	6.51	1.00	0.89	0.381	-	Amplification
OLFR198	1	6.57	6.57	1.00	-1.45	0.157	-	Amplification
LSM12	1	7.06	7.06	1.00	1.63	0.113	-	Amplification
CCDC115	1	6.96	6.96	1.00	-2.49	0.018	-	Amplification
MED25	1	8.91	8.91	1.00	-3.00	0.005	-	Amplification
DNAJB12	1	7.43	7.43	1.00	-4.35	1.38E-04	-	Amplification
CLCF1	1	6.61	6.61	1.00	-4.34	1.42E-04	-	Amplification
HIST1H1B	2	6.70	6.70	1.00	4.15	2.39E-04	-	Amplification